

Cluster Dynamics with HETDEX - I: Simulated Performance, Mass Distribution and Limits

Steven Boada,¹★ C. Papovich,¹ R. Wechsler,^{2,3} T. S. Li,¹
K. Gebhardt,⁴ E. Rozo,² E. S. Rykoff,²

¹George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX, 77843-4242

²Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, Stanford, CA 94305, USA

³Department of Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

⁴The University of Texas

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The study of clusters of galaxies has been argued to be a very effective way to measure cosmological parameters, including measuring dark energy and testing models of gravity. The Hobby Eberly Telescope Dark Energy Experiment (HETDEX) will observe many hundreds of square degrees, covering a large sample of galaxy clusters out to $z = 0.5$ based on their optical spectra (3500 – 5500Å). The spectra will provide important measures of the clusters dynamics and may enable constraints on cosmological parameters, but only if the measurements provide accurate estimates of the total cluster masses. We have carried out a study to investigate the ability of HETDEX to recover accurate galaxy cluster masses over a wide range of masses and redshifts. We used a detailed mock galaxy catalog and present mock observations of two different scenarios: (1) We targeted individual galaxy clusters to investigate the recovery of parameters with such observations. (2) We created and evaluated a HETDEX-like selection “function” of galaxies over a similarly sized portion of the sky and use well adopted techniques to recover the dynamical properties, such as velocity dispersion and mass. Using both observing strategies, we produce cluster mass probability density functions $P(X|M, z)$, which can be used to determine the probability that a galaxy cluster of given mass (M), located at redshift (z) determined using observable parameter (X).

1 INTRODUCTION

Our ability to perform precision cosmology with clusters of galaxies has reached a critical turning point. The widely accepted Λ CDM model of cosmology makes explicit predictions about the number and masses of galaxy clusters throughout the universe. However, connecting these predictions to a set of, sufficiently large in size, observed clusters remains a principal problem. Specifically, the largest threat to modern, precision, cluster cosmology is not the identification of large numbers of clusters (the total number of clusters known is only going up) but the accurate recovery of galaxy cluster mass (e.g., Sehgal et al. 2011; Planck Collaboration 2013; Bocquet et al. 2015).

As mass is not a direct observable, a lot of work is underway to characterize galaxy cluster masses with an observable feature of galaxy clusters. The goal is to constrain, as best possible, $P(X|M, z)$ the probability (P) that a galaxy cluster of given mass (M), located at redshift (z) determined using observable parameter (X). The observable parameter most commonly observed X-ray temperatures and luminosities (e.g., Mantz et al. 2010; Rykoff et al. 2014; Mantz et al. 2015), cosmic microwave background observa-

tions (e.g., Vanderlinde et al. 2010; Sehgal et al. 2011) using the Sunyaev-Zel’dovich effect (SZE; Sunyaev & Zeldovich 1972) or optical studies (e.g., Rozo et al. 2010, 2015) of richness (e.g., Abell 1958; Rykoff et al. 2012) or galaxy velocity dispersions (e.g., Ruel et al. 2014; Sifón et al. 2015b).

Massive surveys, both on going and planned, are revolutionizing cluster cosmology using a large range of wavelengths. The South Pole Telescope (SPT; Carlstrom et al. 2011) and the Atacama Cosmology Telescope (ACT; Swetz et al. 2011) are discovering many clusters through the SZE. Optically, the on going The Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) and planned Large Synoptic Survey Telescope (LSST; LSST Dark Energy Science Collaboration 2012) will identify many thousands of clusters to much lower masses than is possible with SZE measurements. However, regardless of the discovery method used, spectroscopic followup is needed to further constrain $P(X|M, z)$. But as the cluster dataset grows to many tens of thousands of clusters individual followup becomes increasingly impractical. Therefore, large spectroscopic surveys are needed to more fully constrain the observable–mass relation of clusters.

The Hobby Eberly Telescope Dark Energy eXperiment

(HETDEX; Hill et al. 2008) is a trailblazing effort to observe high-redshift large scale structures using cutting edge wide-field integral field unit (IFU) spectrographs. Designed to probe the evolution of the dark energy equation of state etched onto high redshift ($z > 2$) galaxies by the Baryon Acoustic Oscillations (Eisenstein et al. 2005) in the first moments of the universe, the survey will observe two fields for a total of 420 degree² from two fields (300 degree², Spring field and 120 degree², Fall field). Tuned to find Ly α emitting (LAE) galaxies at $1.9 < z < 3.5$, HETDEX expects to find 800,000 LAEs, and more than one million [O II] emitting galaxies at $z < 0.5$ masquerading as high-redshift galaxies (Acquaviva et al. 2014).

While a large portion of the $\sim 10^6$ interloping [O II] galaxies will be field (not associated with a bound structure) galaxies, the large area covered by HETDEX is expected to contain as many as 100 Virgo-sized ($M_{\text{dyn}} \sim 10^{15} M_{\odot}$) clusters at $z < 0.5$ (citation?). The near-complete spectroscopic coverage allows an unprecedentedly detailed look at a very large number of clusters ranging from group scales to the very massive. In addition to the recovery of accurate dynamical masses, detailed investigations of the of dynamical state of the clusters is possible.

Connecting the dynamical properties derived from spectroscopy to the properties inferred from other studies insures the greatest impact on future work. HETDEX overlaps with the Sloan Digital Sky Survey (SDSS; Blanton et al. 2001), SDSS stripe 82 (Annis et al. 2014), the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005), and the upcoming DECam/IRAC Galaxy Environment Survey (DIRGES; PI: Papovich, C. Papovich et al. in preparation). **SHELA and others? Would be good to have a whole list of different things and different wavelengths.** While the potential dataset is very rich, two large issues remain.

It is unclear how a blind spectroscopic survey with an IFU will effect the recovery of galaxy cluster dynamical properties. Unlike many previous large cluster surveys (e.g., Milvang-Jensen et al. 2008; Robotham et al. 2011; Sifón et al. 2015b) which use multi-object spectrographs, the Visible Integral-Field Replicable Unit Spectrograph (VIRUS; Hill et al. 2012) used by HETDEX samples the sky unevenly which could excluded member galaxies which would otherwise be included. Secondly, it is not straightforward to use spectroscopic redshifts predominately from emission-line galaxies to interpret the kinematic and dynamical states of the clusters.

The wording here needs to be updated when the time comes. This work plans to address these concerns in the following ways. We use simulated observations which target individual galaxy clusters to investigate the recovery of parameters with such observations. Secondly, we create and evaluate a HETDEX like selection “function” of galaxies over a similarly large portion of the sky and use well adopted techniques to recover the dynamical properties, such as velocity dispersion and mass. Each observation strategy will further be constrained with “ideal” and “realistic” knowledge. Ideal knowledge assumes that we know which individual galaxy is assigned to which cluster. With realistic knowledge this is unknown and must be determined prior to the estimation of the cluster properties. Both of these strategies will better allow future work to predict the number and types of galaxy clusters which should be observed with VIRUS during both

the HETDEX survey portion and through targeted follow up observations.

We begin in Section 2 by giving an overview of what data is used, how it is created, and how we make our “observations.” Details about the determination of cluster parameters, velocity dispersion, total mass, etc., are discussed in Section 3. Next, we present the results of our study in Section 4 and discuss their implications in Section 5. Finally we summarize our findings in Section 6.

Throughout this paper, we adopt the following cosmological model ($\Omega_{\Lambda} = 0.77$, $\Omega_M = 0.23$, $\sigma_8 = 0.83$ and $H_0 = 72 \text{ km s}^{-1} \text{ Mpc}^{-1}$), assume a Chabrier initial mass function (IMF; Chabrier 2003), and use AB magnitudes (Oke 1974).

2 DATA AND MOCK OBSERVATIONS

In this section, we describe the data products and the techniques used to replicate the HETDEX survey. We use the information from a large mock galaxy catalog supplemented by galaxies from the SDSS to create a realistic “sky” and “observe” it with a HETDEX-like observing strategy.

2.1 The “Buzzard” Catalogs

The “Buzzard” mock galaxy catalogs (R. Wechsler et al., private communication) cover 375.68 degree² between $60 < RA < 90$ and $-61 < DEC < -41$ and are derived from a combination of Sub-halo Abundance Matching (ShAM) and ADDSEDs (Adding Density Dependent Spectral Energy Distributions) tied to an in house n-body cosmological simulation. A brief description of the catalog creation is described as follows. The initial conditions are generated with a second-order Lagrangian perturbation theory using 2LP-Tic (Crocce et al. 2006). Dark matter (DM) n-body simulations are run using LGadget-2 (a version of Gadget-2; Springel 2005). The DM halos are identified using the ROCKSTAR halo finder (Behroozi et al. 2013) which also calculates halo masses and other various parameters.

Galaxy M_r luminosities are added to the velocity peaks using ShAM (Reddick et al. 2013), and ADDSEDs (Adding Density Dependent Spectral Energy Distributions) assign luminosities in the other bands. A M_r -density-SED relation is created using a SDSS training set, and for each mock galaxy the SED of a randomly selected training set galaxy which has a similar M_r and density is assigned. The result is a 398.49 sq. degree mock catalog occupying a $60 \leq RA \leq 90$ and $-40 \leq DEC \leq -61$ portion of the sky. It contains 238 million galaxies with $r \text{ mag} < 29$ and $z \leq 8.7$.

The catalog information, used in this study, is broken into two large portions. The “truth” files contain the characteristics of each individual galaxies, such as right ascension (RA), declination (DEC), redshift (z), observed and rest-frame magnitudes, and many others. The “halo” files contain information for individual halos, to which many individual galaxies may belong. This includes five estimations of dynamical mass, RA, DEC, z , three dimensional velocity dispersion, and many others. However, the catalogs do not include information for emission lines. We supplement the catalogs by generating this information; the process is described in Section 2.2.

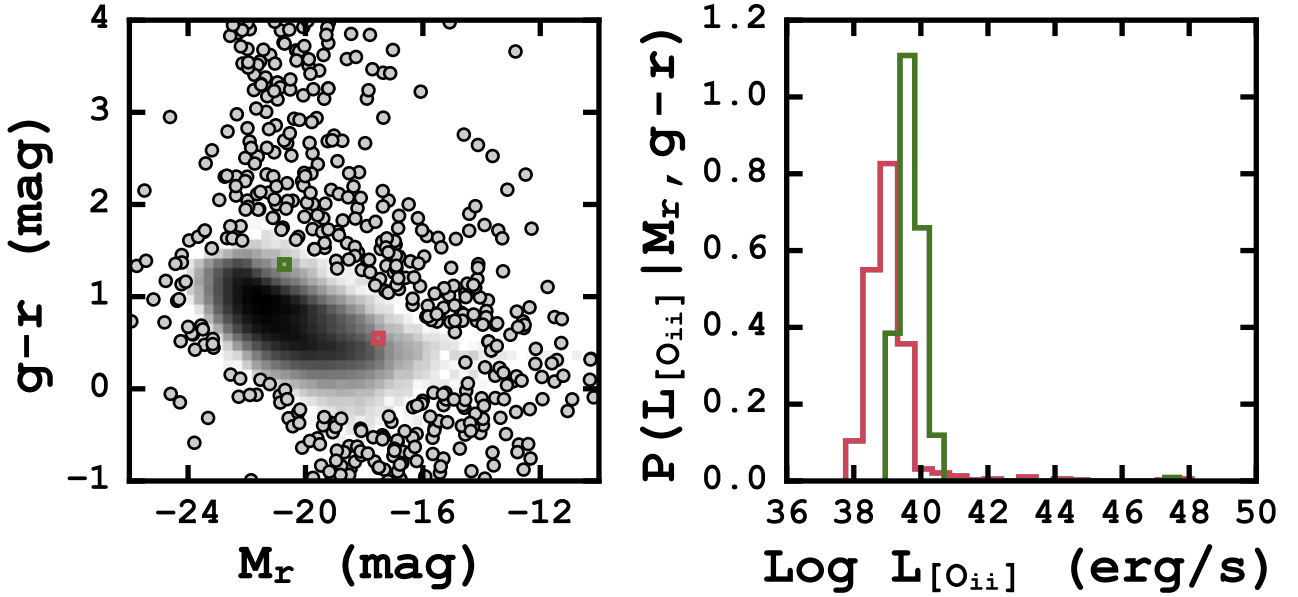


Figure 1. *Left:* CMD of 503113 $z < 0.2$ galaxies taken from the SDSS DR12 where the shading scales with the density of points. The two colored boxes show regions containing potential catalog galaxies. *Right:* Probability histograms of the Log [O II] luminosity for the SDSS galaxies located in the two highlighted regions on the right. New [O II] luminosity (and subsequently fluxes) are assigned to catalog galaxies from slice sampling the probability histogram. See the text for a full description of the process.

We investigate the accuracy of the halo mass distribution by comparing the cumulative number density of halos above a mass (M_{200c}) threshold to the halo mass function (HMF) of Tinker et al. (2008). We calculate the HMF is calculated at central redshifts of 0.1, 0.2, and 0.4 using HM-Fcalc (Murray et al. 2013) and compared to galaxies in a redshift window of $\Delta z \pm 0.01$. We find a very good agreement between the predicted HMF and the observed distribution of clusters.

2.2 [O II] Luminosity

The Buzzard “truth” catalog does not provide [O II] luminosities so we must assign them empirically. We use 503113 galaxies from the SDSS Data Release 12 (Alam et al. 2015) from $z = 0.05 - 0.2$, which are selected with no redshift warning, and place each galaxy on a color-magnitude diagram (CMD) of M_r and $g - r$; see the left panel of Figure 1.

To assign an [O II] luminosity to each galaxy in our catalog we place the catalog galaxies on the same CMD and select all SDSS galaxies in a small 2D ($M_r, g - r$) bin around the galaxy. We extract all of the SDSS galaxies inside that bin and create a histogram of their [O II] luminosities. Using a slice sampling technique (Neal 1997) we assign the catalog galaxy an [O II] luminosity based on the distribution of SDSS galaxies extracted. For catalog galaxies which are placed on the CMD near no, or very few ($1 \leq n < 10$) galaxies we assign it zero [O II] luminosity or the mean luminosity, respectively.

The right panel of Figure 1 shows the CMD of all SDSS galaxies. Two potential catalog galaxies are also placed on the CMD ($M_r, g - r = -17.7, 0.49$ and $M_r, g - r = -21.4, 1.24$) and indicated by the two colored boxes. The histograms show in the Figure’s right panel show the probability density

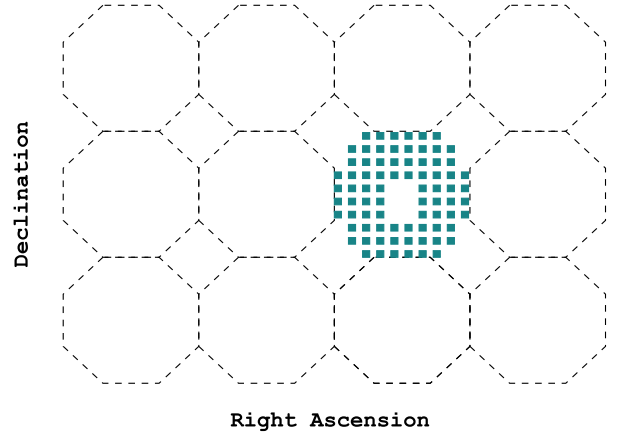


Figure 2. Representative observation tiling scheme for the HETDEX $16' \times 16'$ pointings. Each colored square is a single VIRUS IFU and the dashed octagons approximate the size of a single observation. See the text for more details.

histograms of the Log [O II] luminosity for the SDSS galaxies in the 2D bin around the individual catalog galaxies. We sample the distribution and assign each catalog galaxy an [O II] luminosity which is then converted into a flux.

2.3 Mock Observations

Tentatively slated to start in the spring of 2016, HETDEX will perform blind spectroscopy ($R \sim 750$ in $3500 - 5500 \text{ \AA}$) over two fields along the celestial equator. The 300 degree^2 ,

spring field and 120 degree², fall field will have no preselected targets. Using VIRUS on the 10-m Hobby-Eberly Telescope (HET; Ramsey et al. 1998) the completed survey is expected to have an overall fill-factor of 1/4.5, meaning that the entire area could be covered with 4.5 dithers of the entire survey.

Each mask is created to accurately reproduce the HETDEX IFU pattern, see Figure 2. The pattern consists of 78 IFUs, which are comprised of 448 optical fibers subtending a 50''×50'' region on the sky (Kelz et al. 2014). The inter-IFU spacing is also 50'' spanning a total area of 16'×16' on the sky.

The individual IFUs have a fill-factor of 1/3, which will be completely filled with three dithers of the telescope at each pointing. This means that when selecting galaxies from the Buzzard catalog we assume an observation for all galaxies laying within a colored, IFU square in Figure 2. Galaxies which lie between the IFUs are missed, as well as the galaxies which lie between the pointings, as there is no overlap between one pointing and the next. To cover the 375.67 degree² field of the Buzzard catalog we require 5370 pointings where 0.015 degree² of each pointing is covered by an IFU. The total area of the sky covered by an IFU is 80.80 degree² which gives a filling factor of 1/4.65 slight decreased from the expected filling factor of 1/4.5.

The spectral coverage allows for the detection of [O II] ($\lambda\lambda 3727 - 3729$ Å doublet) emitters to $z \sim 0.5$ and Ca H ($\lambda 3968.5$ Å) and K ($\lambda 3933.7$ Å) absorption features to $z \sim 0.4$. HETDEX is expected to detect sources with continuum brighter than $g = 22$ mag, and emission line strengths above 3.5×10^{-17} erg s⁻¹ cm⁻².

In this work we consider two separate observing strategies and two detection limits. The two observing strategies are targeted and survey-like. The targeted observations use “direct” observations where each cluster is targeted individual, and every cluster member galaxy is assumed to be observed. The survey observations mimic the HETDEX observation pattern across the sky, where no cluster is directly targeted and not all cluster member galaxies are observed. In both cases we use two detection thresholds. “Perfect” observations assume no detection threshold, if a cluster member galaxy is observed, it is also detected. “Realistic” observations assume the HETDEX-like detection thresholds where a galaxy may be observed but not necessarily detected.

These observations provide four levels of quality with targeted-perfect being the highest and survey-realistic being the lowest.

3 RECOVERY OF PARAMETERS

In the following sections, we outline the methods we use to derive the dynamical properties of the galaxy clusters in our sample. This is not meant to be an exhaustive study of the different methods used to recover these parameters. The following is, in many cases, a subset of the available methods to derive any single parameter. The specific choice of method may improve or diminish the accuracy of the recovered parameter, but the methods chosen were to facilitate comparison with observational studies.

3.1 Cluster Redshift

The accurate determination of the cluster redshift (z_c) is crucial to the reliability of all following measurements. An incorrect cluster redshift introduces errors into the measured line-of-sight velocity (LOS_V) and corresponding dispersion, which, in turn, contributes to errors associated with dynamical mass and radius.

In simple terms, the cluster redshift is the mean of the redshifts of all galaxies associated with the cluster. However, because the standard mean can be quite sensitive to outliers or otherwise contaminated data, we require a more resistant statistic, and turn to the biweight location estimator (Beers et al. 1990) which provides improved performance.

3.2 Line-of-Sight Velocity Dispersion

We first calculate the line-of-sight velocity (LOS_V) to each galaxy, where

$$LOS_V = c \frac{z - z_c}{1 + z_c} \quad (1)$$

and c is the speed of light in km s⁻¹, z is the redshift of the individual galaxy, and z_c is the overall cluster redshift described in the previous section.

The line-of-sight velocity dispersion (LOS_VD) is calculated using a method of maximum likelihood following Walker et al. (2006). We maximize the probability function

$$p(\{v_1, \dots, v_N\}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi(\sigma_p^2 + \sigma_i^2)}} \exp\left[-\frac{1}{2} \frac{(v_i - \langle\mu\rangle)^2}{(\sigma_p^2 + \sigma_i^2)}\right] \quad (2)$$

where σ_p , $\langle\mu\rangle$, and σ_i is the LOS_VD, the average radial velocity and the error on the individual LOS_Vs respectively. Using a Monte Carlo Markov Chain (MCMC) sampler (**emcee**¹; Foreman-Mackey et al. 2013) which is based on affine-invariant ensemble sampler (see Goodman & Weare 2010 for details on affine-invariant samplers). We draw twenty thousand samples from the posterior probability distribution. Simple priors, $\langle\mu\rangle$ lies between the maximum and minimum LOS_V and $0 < \sigma_p < 1400$ km s⁻¹, are used. When the full distribution of LOS_VDs are not used, the final LOS_VD is quoted as the median value of the posterior probability distribution with 68% error bars defined as the 16th and 84th percentiles of the same distribution.

In principle, a single statistic such as the biweight scale estimator or the gapper estimator (both from Beers et al. 1990) with many bootstrap resamplings could be used to construct a distribution of σ_p . In simple tests where the values of both σ_p and $\langle\mu\rangle$ are known. The 68% error bars derived from the MCMC method give slightly better results with the true LOS_VD value bracketed by the error bars in $\sim 68\%$ of the cases versus $\sim 57\%$ with bootstrapping and a single statistic. In addition, we prefer the maximum likelihood method for its straight forward treatment of the errors in the LOS_V measurements.

¹ <http://dan.iel.fm/emcee/current/>

3.3 Dynamical Mass

Recently, the relationship between the LOSVD and dynamical mass has been the focus of several studies (e.g., [Evrard et al. 2008](#); [Saro et al. 2013](#); [Sifón et al. 2013](#); [van der Burg et al. 2014](#)), and a best fitting relationship for the mass enclosed by r_{200c} of the form

$$M_{200c} = \frac{10^{15}}{h(z)} \left(\frac{\sigma_{1D}}{A_{1D}} \right)^{1/\alpha} M_{\odot} \quad (3)$$

with $A_{1D} = 1177 \pm 4.2 \text{ km s}^{-1}$ ([Munari et al. 2013](#); referred to as σ_{15} in [Evrard et al. 2008](#) and other works), $\alpha = 1/3$, $h(z) = H(z)/100$, and σ_{1D} is the LOSVD of the velocity tracers (dark matter particles, subhalos or galaxies).

A growing body of work suggests that there is a significant difference in the observed LOSVD depending on the velocity tracers used. Specifically, while there is little difference between using galaxies and their host DM subhalos, there is a significant over estimation of the LOSVD when using galaxies/subhalos compared to DM particles ([Munari et al. 2013](#)). We follow other works (e.g., [Kirk et al. 2015](#); [Sifón et al. 2015a](#)) using the scaling relation, given in Equation 3 from [Munari et al. \(2013\)](#) to facilitate comparisons with other observational studies.

3.4 Dynamical Mass Corrections

In this section we use two methods to predict the mass of a cluster based on other observables. Often the cluster mass is estimated based on a single observable, X-ray temperature, velocity dispersion, richness and others. Here we combine many observables to attempt to correct the mass inferred solely from the velocity dispersion. The first method is traditional probability based where we marginalize over a series of observables to find the most probable mass. The second is based on a machine learning (ML) algorithm which attempts to “learn” the relationship between the observables and the desired output, the mass. Both of these methods are examples of supervised learning algorithms where the relationship between the observable (known) parameters and the target parameter (the mass) are both known.

As with any predictive analysis it is important to test the model on data that the model has not seen before. In this section we take all of the observed clusters, our full sample, split them, and generate a training and testing set. The data is randomly split 70% training and 30% testing. We follow the ML convention and refer to the individual clusters in each set as a “sample”, and the parameters associated with the cluster (z , LOSVD, mass, etc.) as “features”.

3.4.1 Probability Based

We begin with the training sample. After selecting the desired features $\vec{x} = \{\sigma, z, \dots\}$ we make the joint probability between the true cluster mass (M) and \vec{x} . Because \vec{x} can be multidimensional, we rely on the corner plot to visualize the relationship between all of the training features. Figure 3 shows all of the one (marginalized probability) and two (joint probability) dimensional projections of the posterior probability distributions of the features of the training data.

The conditional probability of the mass $P(M|\vec{x}) =$

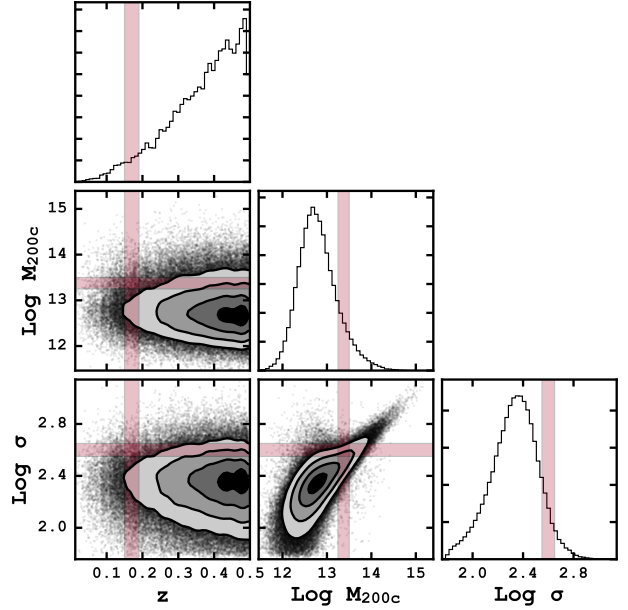


Figure 3. Corner plot of the *training* data with features σ and z . The corner plots shows all of the one and two dimensional posterior probability distributions used to determine the correct cluster mass. The colored rectangles, which have been enlarged for clarity, show the slices needed to create a conditional probability distribution of the mass, $P(M|\vec{x})$. See text for a complete description.

$\{x_1, x_2, \dots\}$) is determined by taking a slice through the joint probability distributions in bins centered on the desired value. The slices shown by the colored bars in Figure 3 are centered on $\sigma = 400 \text{ km s}^{-1}$ and $z = 0.17$. The distribution of mass contained in the three dimensional bin given by the intersection of these slices is $P(M|\vec{x} = \{\sigma = 400 \text{ km s}^{-1}, z = 0.17\})$.

For the clusters making up the *test* sample the mass is unknown (it is what we are trying to predict) but the other features are known. To determine the mass probability distribution of a test cluster, $P(M)$ we combine the conditional probability distribution, $P(M|\vec{x})$, created previously with the probability distribution of σ through Equation 4.

$$P(M) = \int P(M|\vec{x})P(\sigma)d\sigma \quad (4)$$

The expected mass is determined by integrating the mass probability, $P(M)$ over all mass. This becomes our “predicted” mass, $\langle M \rangle$.

$$\langle M \rangle = \int M' P(M') dM' \quad (5)$$

The confidence interval associated with this prediction can be estimated two ways. First, by calculating the variance about the expected mass through

$$V = \int (M' - \langle M \rangle)^2 P(M') dM' \quad (6)$$

or by drawing many samples from $P(M)$ and calculating the values at the 16th and 84th percentile. In practice we find that both methods produce similar results for a large number

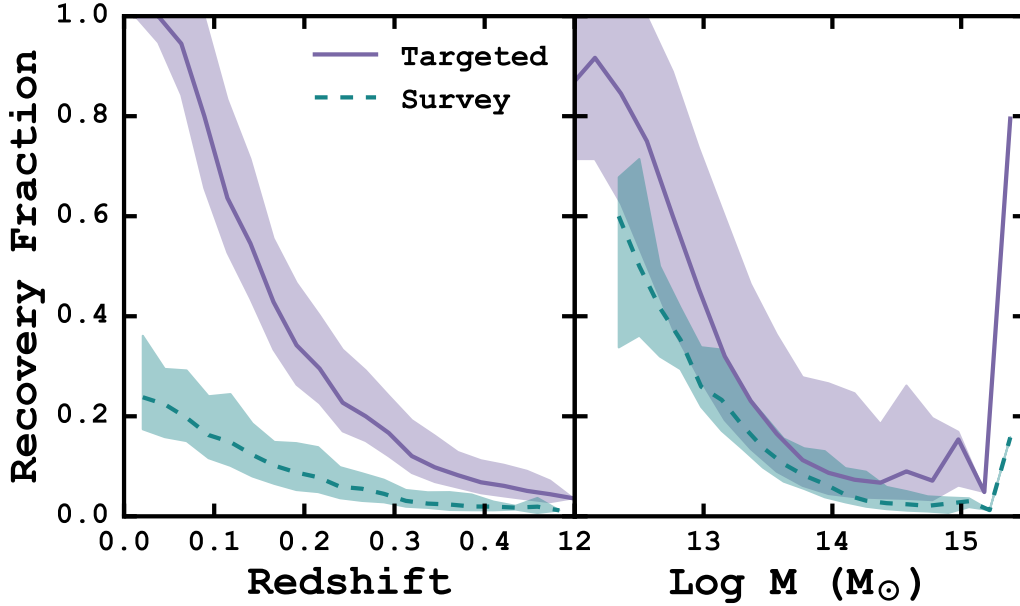


Figure 4. Recovery fractions (N_{obs}/N_{true}) of cluster member galaxies as a function of redshift and mass for the targeted and survey observing strategies. The solid lines are the median values and the shaded regions represent the 68% scatter. The significant decline in galaxies observed with the survey strategy is due to gaps in the VIRUS IFU.

of trials. Therefore, we quote predicted masses as the most probable mass given by Equation 5 and associated 68% error estimated through Equation 6.

3.4.2 Machine Learning Based

The estimation in this section relies on a ML technique known as an ensemble method, where many estimators are created by a single learning method with the goal of improved generalization and robustness compared to a single estimation. Ensemble methods come in two general flavors. Averaging methods average (hence the name) the estimators to produce a single prediction. Boosting estimators build estimates sequentially by attempting to address poor performing estimators in each previous step, hence “boosting” the predictive power.

Here we use an averaging ensemble learning method known as a forest of randomized decision trees often shortened to just random forest (RF). Decision trees can be visualized as a flow chart where forks are the branches of the tree. The path along the tree is decided by the values of the feature at each branch. RF estimators use a random subset of the training set at each fork to decide which path should be followed. The final prediction is then the average of all the trees. We use RF regression methods as implemented in **Scikit-Learn** (Pedregosa et al. 2012).

Any uncertainties quoted by this method are prediction intervals not confidence intervals. A prediction interval is an estimate of the interval encompassing future observations, with a certain probability. And, unlike confidence intervals, which describe certainties on the different moments of a population, a prediction interval is unique to each prediction. In many regression analyses, such as linear fitting, the prediction intervals are based on underlying assumptions of nor-

mally distributed residuals. However, RF estimators do not have any such assumptions and require special treatment.

The prediction intervals here are based on the general method of quantile regression forests (Meinshausen 2006). The general idea is that all response variables are recorded, not just the mean. Then the prediction can be returned as the full conditional probability distribution of all responses, which allows us to generate the prediction intervals. The 68% prediction interval is determined by calculating the 16th and 84th percentile of the full conditional probability distribution. *I am going to change this to just the std of the distribution. I need the errorbars to be symmetric to make the fitting routine easy later on.*

4 RESULTS

Here we explore the cluster member recovery rate and mass estimates for the two observing strategies. We discuss the accuracy of dynamical mass derived from both the scaling relation (see Equation 3) and through the probability and ML methods.

4.1 Recovery of Cluster Members

As discussed in Section 2.3 the observational constraints place limits on the total number of cluster member galaxies expected to be recovered. Knowing these limits will provide important information for potential future follow up or targeted observations.

Figure 4 shows the recovery fraction of member galaxies, the number of observed galaxies divided by the number of actual galaxies (N_{obs}/N_{true}), as function of both redshift and cluster mass. It is important to note that if few than

five member galaxies are observed the cluster is not considered detected, and is excluded from this figure. As expected, the targeted observing strategy where the individual clusters are targeted through several dithers to ensure near complete coverage, performs significantly better than the survey observing strategy across all redshifts and cluster masses.

For the clusters recovered as a function of redshift, there are two effects at work. The decrease in recovery fraction with increasing redshift is a magnitude effect. If, instead of being limited to $g = 22$ mag, the observations were limited by absolute magnitude, the sharp downward trend disappears. The second key feature is the strong decline in clusters recovered from survey observations. This is due to gaps in the VIRUS IFU. The median recovery fraction in survey observations is almost exactly 4.5 times less than the targeted median recovery fraction. As the total filling factor of the survey increases the two lines will converge.

The recovery rate as a function of cluster mass, right panel of Figure 4, shows that of the low mass clusters we detect ($N_{obs} > 5$), observe the majority of the galaxies. This also shows a rapid decrease in the detection fraction, which can again be explained by considering absolute magnitudes instead of apparent magnitudes. Just as before, if the survey was limited by absolute magnitude, we find a much more consistent detection fraction.

4.2 Mass estimates

In this section we discuss how accurately we are able to reproduce the true cluster mass from a set of observations. We report on two methods the probability based approach (Section 3.4.1) and the ML based method (Section 3.4.2). For each method we consider both targeted and HETDEX-like observing strategies.

In both figures, we include the cluster masses recovered through the power law scaling relation given in Equation 3 for both the targeted and survey observations. It should serve as baseline to compare the probability based and ML cluster mass recovery methods. And, while there are many possible metrics to evaluate performance, we compute two: the median absolute error (MAE) and the root mean squared error (RMSE). Both metrics evaluate how closely the ensemble of predicted cluster masses are to the true cluster masses, and in both cases lower numbers are better.

Just because a predicted cluster mass may not exactly match the true value does not mean it is a bad prediction. In addition to the MAE and RMSE, we also report the fraction of predictions where the true cluster mass is contained within the 68% confidence or prediction intervals (see Sections 3.4.1 and 3.4.2).

The cluster masses predicted by Equation 3 gives the following results. The MAE is 0.263 dex and for both the targeted and survey observations. The RMSE is 0.396 dex and 0.394 dex for the targeted and survey observations respectively. This scatter in recovered masses can be attributed to both physical and numerical effects. As the cluster mass increases, clusters become more virialized and contain many individual galaxies. The presence of any in-falling matter onto lower mass clusters can introduce a significant amount of substructure, which can increase the observed LOSVD increasing the predicted mass. Also, as the number of cluster galaxies decreases the LOSVD PDF is poorly sampled

leading to poorly recovered cluster masses due to numerical effects. The masses presented here are recovered using the best possible conditions, where we have perfect knowledge of the cluster membership. In reality, the mass recovery levels presented in this section represent an upper bound (the best) on the accuracy achievable through this method. [Ntampaka et al. \(2015\)](#) does have a discussion about how well they do with contaminated galaxy catalogs. We could do something similar and have a similar discussion, but I'm not sure it is worth it. Should be simple enough to do with the targeted catalog, but with the HETDEX catalog it would be pretty bad IMO. We should also talk about how often the true mass lies within the error bars. Many of them are going to be with the 68% range, but we can drop the error estimates to 0.5σ if that actually means anything

The MAE and RMSE associated with the ensemble of predictions are summarized in Table 1. As expected, we find that both the probability based and ML based methods outperform the standard power law based methods for both the targeted and survey observations. We also find that the ML based method produces lower ensemble error for both the observation strategies when compared to the probability based method, although the MAE and RMSE are reduced by only ~ 0.01 dex. [here is where we would talk about cross validation on the errors.](#)

In both Figures 5 and 6, we show the predicted versus true cluster masses for each of the two observing strategies. In each panel the solid black line is the 1:1 relationship, the solid colored line is the median recovered mass for the targeted observing, and the colored, dashed line is the median recovered mass for the HETDEX-like observations. The shaded regions are the 68% scatter around the median values (the 16% and 84% quartiles). The lower panels show the fractional cluster mass error:

$$\epsilon = (M_{pred} - M)/M \quad (7)$$

where M_{pred} is the predicted cluster mass and M is the true cluster mass.

In both figures we successively add additional information to further constrain the true cluster masses which subsequently reduces the error associated with the ensemble of predictions (see Table 1). The power law derived masses, using the single (not including the cosmological parameters) free parameter, σ (the LOSVD), over estimates the predicted cluster mass at all masses. Both the probability and ML methods over predict the mass of low mass clusters and under predict the mass of the higher mass clusters. When the redshift information is also added, the amount of this over and under prediction is lessened but the cross over point, remains roughly consistent at $\sim 10^{14} M_{\odot}$. Additionally, the number of galaxies observed, N_{gal} , further reduces the MAE and RMSE on the predictions but also lowers the cross over point to $\sim 10^{13.5} M_{\odot}$.

[Need to talk about the cluster rotations to fill out the survey data. Otherwise there aren't that many clusters detected.](#)

Table 1. Summary of the errors associated with the Targeted and Survey observation strategies, as an ensemble of predictions. See the text for discussion about the MAE and RMSE. Overlap is the percentage of clusters where the true cluster mass is bracketed by the prediction intervals of the predicted mass. See Sections 3.4.1 and 3.4.2 for a discussion on prediction intervals.

	Input Features	Targeted			Survey		
		MAE (dex)	RMSE (dex)	Overlap (%)	MAE (dex)	RMSE (dex)	Overlap (%)
Prob Based	Power Law	0.259	0.394	...	0.243	0.357	...
	σ	0.220	0.337	...	0.210	0.310	...
	σ, z	0.193	0.283	...	0.184	0.271	...
	σ, z, N_{gal}	0.128	0.204	...	0.141	0.215	...
ML Based	Power Law	0.258	0.394	...	0.255	0.380	...
	σ	0.241	0.359	...	0.223	0.342	...
	σ, z	0.193	0.282	...	0.197	0.290	...
	σ, z, N_{gal}	0.118	0.192	...	0.140	0.221	...

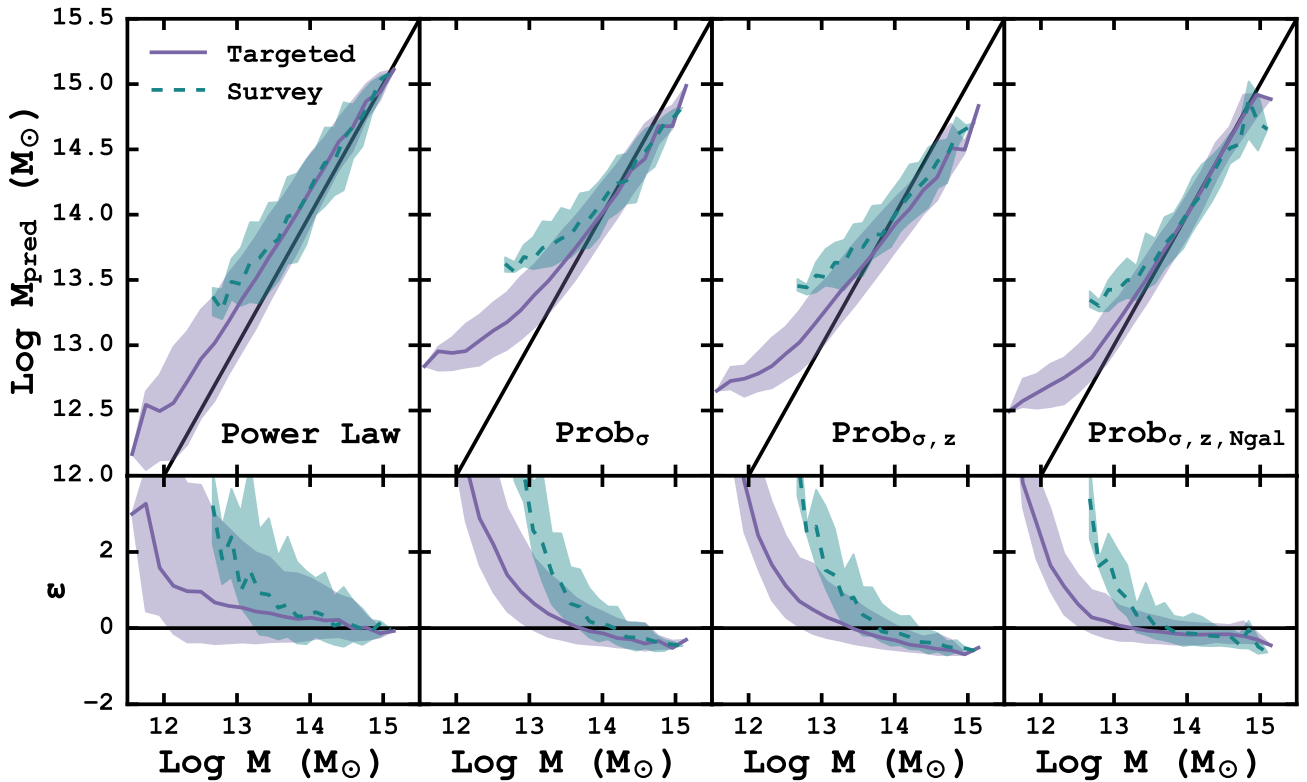


Figure 5. Mass predictions for the power law scaling relation (Equation 3) and the probability based technique with different input features as a function of true cluster mass. The bottom row of panels shows the fractional error (Equation 7) also as a function of true cluster mass. The solid black line shows the 1:1 relation. The solid, colored line is the median predicted mass for the targeted observing, and the colored, dashed line is the median recovered mass for the HETDEX-like observations. The shaded regions represent the 68% scatter around the median values.

5 HETDEX AS A GALAXY CLUSTER SURVEY

5.1 Extendability to Other Surveys

Large-scale optical surveys (e.g., DES and LSST) expect to detect hundreds of thousands of galaxy clusters at $z < 1$. Because they are photometric, a major challenge for these surveys is relating a cluster observable to the total dark-matter mass. One promising mass estimator is the optical

richness (e.g., Abell 1958). Specifically, here, we use λ , the weighted number of galaxies within a scale aperture (e.g., Rozo et al. 2011) as calculated by the redMapper algorithm (Rykoff et al. 2012). Previous works (e.g., Rozo et al. 2010) show that the richness correlates strongly with cluster mass on the average, but the absolute mass scale of the optical richness mass estimator and the scatter in cluster mass at fixed optical richness are imprecisely known (Rykoff et al. 2012). These systematics remain the major source of un-

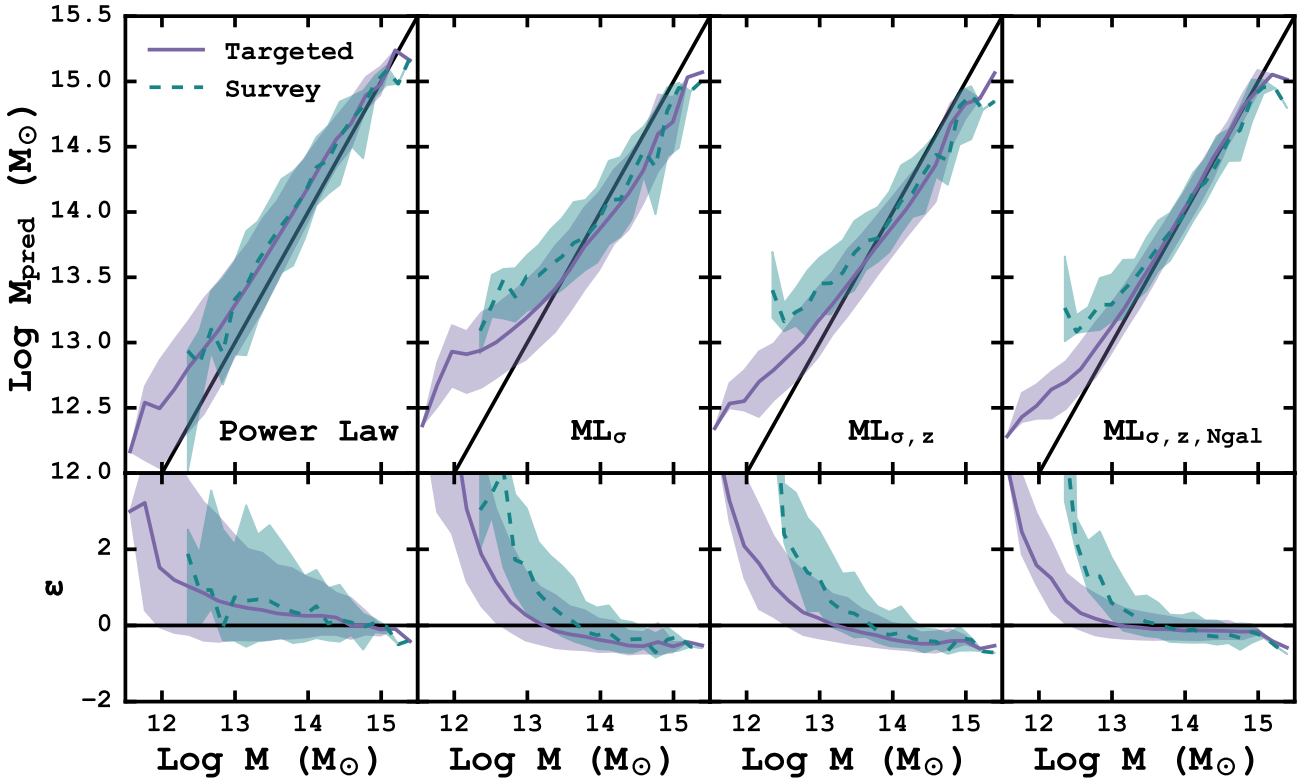


Figure 6. Mass predictions for the power law scaling relation (Equation 3) and the ML based technique with different input features as a function of true cluster mass. The bottom row of panels shows the fractional error (Equation 7) also as a function of true cluster mass. The solid black line shows the 1:1 relation. The solid, colored line is the median predicted mass for the targeted observing, and the colored, dashed line is the median recovered mass for the HETDEX-like observations. The shaded regions represent the 68% scatter around the median values.

Table 2. Associated mass prediction errors (MAE) for different bins of predicted mass. This table shows the relative error for the targeted (top section) and survey (bottom section) observations for different predicted mass bins. The values in the table correspond to the MAE for the galaxies in the predicted mass bin (top row) for the different mass recovery strategies (leftmost column). It can be used to understand how precise the predicted cluster mass is.

Input Features		Bins – $\text{Log } M_{\text{pred}} (M_{\odot})$							
		11.5 – 12	12 – 12.5	12.5 – 13	13 – 13.5	13.5 – 14	14 – 14.5	14.5 – 15	15 – 15.5
Prob Based	Power Law	0.602	0.376	0.340	0.278	0.225	0.177	0.114	0.041
	σ	1.030	0.655	0.389	0.212	0.160	0.158	0.224	0.328
	σ, z	0.804	0.476	0.242	0.173	0.170	0.210	0.353	0.529
	$\sigma, z, N_{\text{gal}}$	0.649	0.372	0.163	0.124	0.113	0.111	0.105	0.241
ML Based	Power Law	0.602	0.376	0.340	0.278	0.225	0.177	0.114	0.041
	σ	0.773	0.557	0.287	0.210	0.230	0.267	0.337	0.323
	σ, z	0.631	0.380	0.211	0.172	0.182	0.236	0.284	0.265
	$\sigma, z, N_{\text{gal}}$	0.539	0.304	0.148	0.117	0.105	0.095	0.078	0.110
Prob Based	Power Law	0.520	0.319	0.263	0.214	0.140	0.069
	σ	0.765	0.434	0.198	0.149	0.163	0.346
	σ, z	0.592	0.305	0.162	0.165	0.257	0.396
	$\sigma, z, N_{\text{gal}}$	0.476	0.171	0.125	0.141	0.141	15.002
ML Based	Power Law	...	0.583	0.316	0.337	0.278	0.214	0.140	0.069
	σ	...	0.606	0.551	0.263	0.201	0.227	0.200	0.192
	σ, z	...	0.916	0.434	0.206	0.171	0.205	0.249	0.202
	$\sigma, z, N_{\text{gal}}$...	0.780	0.401	0.155	0.129	0.136	0.145	0.142

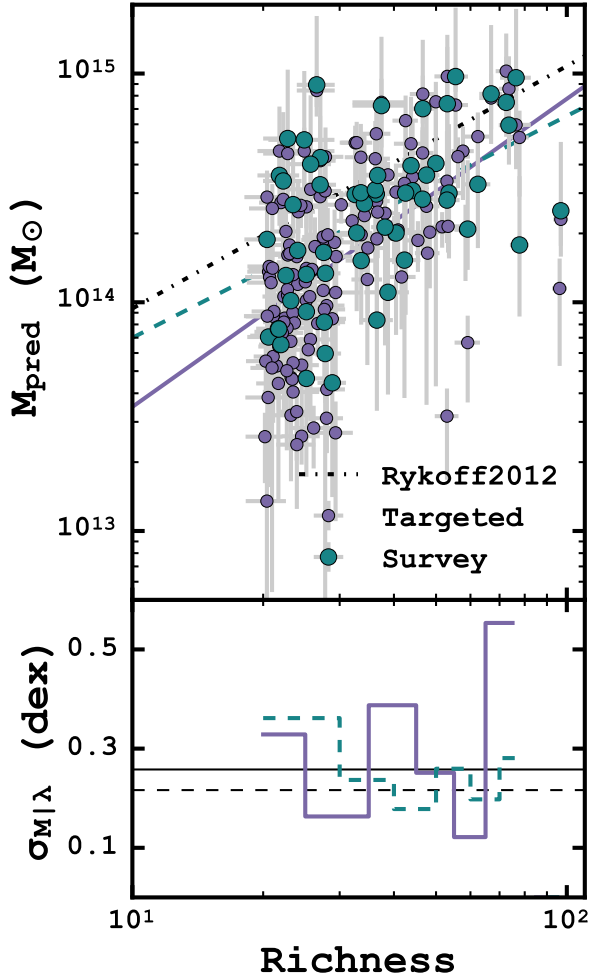


Figure 7. The optical richness, λ , versus the predicted cluster mass. The purple and blue points represent clusters with targeted and survey detections respectively. Error bars show 1σ prediction interval. The solid and dashed lines are fits to either the targeted or survey sets.

certainty in deriving cosmological constraints from cluster abundances and must be measured using independent methods to realize the full potential of these surveys.

We leverage the large spectroscopic dataset of HETDEX to estimate its ability to constrain the scatter and absolute scale of the richness-mass relationship. We begin by cross matching both the targeted and survey observations with a redMapper catalog. The redMapper catalog is limited to $\lambda \geq 20$, or at least twenty object assigned to each cluster. With the targeted and survey observations, we observe 138 and 58 clusters respectively.

Figure 7 shows the optical richness, λ , versus the predicted cluster mass. The cluster masses are the $ML_{\sigma, z, N_{gal}}$ based and correspond to the correct observation strategy. The error bars represent the 1σ prediction intervals for each cluster mass. The solid and dashed lines are fits to the targeted and survey datasets. For comparison, the richness-mass relations from Rykoff et al. (2012) (their equation B5) is shown as a dash-dotted line.

To generate the best fitting lines we follow the general procedure of Hogg et al. (2010), by defining an objective function and then minimizing the loss. Our loss function is

$$\ln L = -\frac{1}{2} \left(\sum_{i=1}^N \frac{[y_i - mx_i - b]^2}{\sigma_{yi}} + \sum_{i=1}^N \frac{[y_i - mx_i - b]^2}{\sigma_{xi}} \right) \quad (8)$$

where we have taken into account the uncertainties in both the richness and predicted cluster mass. We again rely on MCMC samples to sample the posterior probability distribution. The best fitting slope and intercept are quoted as the median value of the posterior probability distribution with 68% error bars defined as the 16th and 84th percentiles of the same distribution.

Following the notation of Rykoff et al. (2012) we find a best-fitting relation for the targeted observations as

$$\ln \left(\frac{M_{200c}}{h_{70}^{-1} 10^{14} M_{\odot}} \right) = 1 \pm 0.13 + 1.34 \pm 0.17 \ln \left(\frac{\lambda}{60} \right)$$

and the survey observations as

$$\ln \left(\frac{M_{200c}}{h_{70}^{-1} 10^{14} M_{\odot}} \right) = 1.02 \pm 0.13 + 0.96 \pm 0.22 \ln \left(\frac{\lambda}{60} \right)$$

The bottom panel of Figure 7 shows the scatter in cluster masses at fixed richness, $\sigma_{M|\lambda}$. The solid and dashed lines represent the targeted and survey observations respectively, and have been offset from one another for clarity. The cluster masses are binned in increasing ten richness intervals (20–30, 30–40, etc.). The two horizontal lines show the mean scatter of 0.26 dex for the targeted observations and 0.21 dex for the survey observations. While not shown in the figure, we find a mean scatter of 0.27 dex if we replace M_{pred} with the true cluster mass for the clusters identified with targeted observations. The reduction in scatter from the targeted to survey observations is due to fewer objects detected in survey mode.

5.2 Potential Improvements

The potential improvements are really three fold. We know that the observing won't cover all of the sky. So we can simply tile more and try to get more clusters. We can observe deeper. I'm not sure the current exposure times (they are in an email that I sent to Casey) but if we can extend them, then we can get further down the IMF and try to build up some of the clusters that are really more like groups than anything else. The last thing we can do is to attempt to build better models to make better predictions on the masses. That might be something like the study that Acquaviva (2016) did with the stellar metallicities. Or we'll need something else that I haven't thought of just yet. The other idea is to come up with some sort of cheat sheet for HETDEX. Basically, if you detect such and such type of galaxy in the survey model then you should go back and try to follow it up with targeted observations. It might be good to talk about the recovery fraction between the two surveys above. Like... if there are 30 cluster members then we should detect the cluster in both surveys.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6 SUMMARY

Here, we present detailed simulations of the upcoming HETDEX survey's applicability to the detection and total mass measurement of galaxy clusters. We use mock galaxy catalogs, and simulated HETDEX-like observational strategies to estimate the number of clusters observed and the precision of their total mass estimates, using three different mass estimates. With an eye to galaxy cluster studies, we discuss potential improvements to the HETDEX survey, and comment on how HETDEX may improve current and future photometric large-area sky surveys' cluster mass estimates derived from optical richness.

Our main conclusions are the following:

(i) HETDEX-like observations do a good job observing galaxy clusters. The number of galaxy clusters detected in the HETDEX-like observations is almost exact 4.5 times less than in the targeted observations. This difference is the same as the overall HETDEX fill factor.

(ii) Overall, we find that a traditional power-law conversion from LOSVD to cluster mass performs significantly poorer than either the probability based or ML based methods also tested. We find the MAE and RMSE of survey detected cluster mass estimates to be approximately 38% and 65% respectively, compared to approximately 123% when a power-law is used alone.

(iii) While the power-law may be outperformed, overall, comparisons of similar mass galaxy clusters shows that above $\text{Log } M = 14.5 M_{\odot}$, the power-law gives a lower MAE than either the probability based or ML based methods. For HETDEX-like observations and clusters with $13 < \text{Log } M < 14.5 M_{\odot}$, we find the $ML_{\sigma,z,N_{gal}}$ method performs the best. Below $\text{Log } M = 13 M_{\odot}$ no method with survey observations gives an MAE of less than 50%.

(iv) For followup targeted observations only galaxy clusters with masses inferred to be $\text{Log } M < 13 M_{\odot}$ from survey observations should be targeted. Galaxy clusters with inferred masses below $\text{Log } M = 13 M_{\odot}$ can see as much as an 81% improvement in cluster mass estimation with targeted observations. While the improvement of clusters with masses above $\text{Log } M = 13 M_{\odot}$ is often approximately 10%.

(v) **Something about the richness.**

It is the author's hope that this work may be useful to others when conducting their own research. Because this work relies heavily on (often) complex data analysis, and in order to promote transparency, we provide all of the code used to conduct this study at <https://github.com/boada/desCluster>. Regrettably, large file size prevents including the source data with the analysis routines. The authors are happy to provide them, if requested.

ACKNOWLEDGEMENTS

The authors also wish to thank the anonymous referee whose comments and suggestions significantly improved both the quality and clarity of this work. We also thank Steven W. Crawford for many helpful discussions. This research made use of This research made use of the IPython package (Perez & Granger 2007) and matplotlib, a Python library for publication quality graphics (Hunter 2007). Figure 3 is heavily based on Foreman-Mackey et al. (2016). Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. **People to thank: Joe, maybe authorship.**

REFERENCES

- Abell G. O., 1958, *ApJS*, 3, 211
- Acquaviva V., 2016, *MNRAS*, 456, 1618
- Acquaviva V., Gawiser E., Leung A. S., Martin M. R., 2014, *Proc. IAU*, 10, 365
- Alam S., et al., 2015, *ApJS*, 219, 12
- Annis J., et al., 2014, *ApJ*, 794, 120
- Beers T. C., Flynn K., Gebhardt K., 1990, *AJ*, 100, 32
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, *ApJ*, 762, 109
- Blanton M. R., et al., 2001, *AJ*, 121, 2358
- Bocquet S., et al., 2015, *ApJ*, 799, 214
- Carlstrom J. E., et al., 2011, *PASP*, 123, 568
- Chabrier G., 2003, *PASP*, 115, 763
- Crocce M., Pueblas S., Scoccimarro R., 2006, *MNRAS*, 373, 369
- Eisenstein D. J., et al., 2005, *ApJ*, 633, 560
- Evrard A. E., et al., 2008, *ApJ*, 672, 122
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Foreman-Mackey D., et al., 2016,] 10.5281/zenodo.45906
- Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
- Hill G. J., et al., 2008, Panor. Views Galaxy Form. Evol. ASP Conf. Ser., 399
- Hill G. J., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Vol. 8446, Ground-based Airborne Instrum. Astron. IV. Proc. SPIE. p. 84460N, doi:10.1117/12.925434, <http://adsabs.harvard.edu/abs/2012SPIE.8446E..0NH>
- Hogg D. W., Bovy J., Lang D., 2010, eprint arXiv:1008.4686
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Kelz A., et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Vol. 9147, Proc. SPIE. p. 914775, doi:10.1117/12.2056384, <http://adsabs.harvard.edu/abs/2014SPIE.9147E..75K>
- Kirk B., et al., 2015, *MNRAS*, 449, 4010
- LSST Dark Energy Science Collaboration 2012, arXiv Prepr. arXiv1211.0310, p. 133
- Mantz A., Allen S. W., Rapetti D., Ebeling H., 2010, *MNRAS*, 406, no
- Mantz A. B., Allen S. W., Morris R. G., Schmidt R. W., von der Linden A., Urban O., 2015, *MNRAS*, 449, 199
- Meinshausen N., 2006, *J. Mach. Learn. Res.*, 7, 983
- Milvang-Jensen B., et al., 2008, *A&A*, 482, 419
- Munari E., Biviano A., Borgani S., Murante G., Fabjan D., 2013, *MNRAS*, 430, 2638

- Murray S., Power C., Robotham A., 2013, *Astron. Comput.*, 3-4, 23
- Neal R. M., 1997, Technical Report 9722, Markov Chain Monte Carlo Methods Based on ‘Slicing’ the Density Function, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.8866&rep=rep1&type=pdf>. Department of Statistics, University of Toronto, Toronto, doi:10.1.1.48.8866, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.8866&rep=rep1&type=pdf>
- Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, eprint arXiv:1509.05409
- Oke J. B., 1974, *ApJS*, 27, 21
- Pedregosa F., et al., 2012, *J. Mach. Learn. Res.*, 12, 2825
- Perez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Planck Collaboration 2013, *A&A*, 571, 19
- Ramsey L. W., et al., 1998, in Proc. SPIE. pp 34–42, doi:10.1117/12.319287, <http://adsabs.harvard.edu/abs/1998SPIE.3352...34Rhttp://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=945060>
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Robotham A. S. G., et al., 2011, *MNRAS*, 416, 2640
- Rozo E., et al., 2010, *ApJ*, 708, 645
- Rozo E., Wu H.-Y., Schmidt F., 2011, *ApJ*, 735, 118
- Rozo E., Rykoff E. S., Bartlett J. G., Melin J.-B., 2015, *MNRAS*, 450, 592
- Ruel J., et al., 2014, *ApJ*, 792, 45
- Rykoff E. S., et al., 2012, *ApJ*, 746, 178
- Rykoff E. S., et al., 2014, *ApJ*, 785, 104
- Saro A., Mohr J. J., Bazin G., Dolag K., 2013, *ApJ*, 772, 47
- Sehgal N., et al., 2011, *ApJ*, 732, 44
- Sifón C., et al., 2013, *ApJ*, 772, 25
- Sifón C., et al., 2015a, eprint arXiv:1512.00910
- Sifón C., Hoekstra H., Cacciato M., Viola M., Köhlinger F., van der Burg R. F. J., Sand D. J., Graham M. L., 2015b, *A&A*, 575, A48
- Springel V., 2005, *MNRAS*, 364, 1105
- Sunyaev R. A., Zeldovich Y. B., 1972, *Comments Astrophys. Sp. Phys.*, 4
- Swetz D. S., et al., 2011, *ApJS*, 194, 41
- The Dark Energy Survey Collaboration 2005, eprint arXiv:astro-ph/0510346, p. 42
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Vanderlinde K., et al., 2010, *ApJ*, 722, 1180
- Walker M. G., Mateo M., Olszewski E. W., Bernstein R., Wang X., Woodroffe M., 2006, *AJ*, 131, 2114
- van der Burg R. F. J., Muzzin A., Hoekstra H., Wilson G., Lidman C., Yee H. K. C., 2014, *A&A*, 561, A79

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.