Cluster Dynamics with HETDEX at z < 0.5 - I: Simulated Performance, Mass Recovery, and Limits to Cosmology

Steven Boada,[1]⋆ C. Papovich,[1] R. Wechsler,[2,3] T. S. Li,[1] K. Gebhardt,[5] J. DeRose,[2] E. Rozo,[2,4] E. S. Rykoff[2]

[1] *George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy,*
Texas A&M University, College Station, TX, 77843-4242, USA
[2] *Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, Stanford, CA 94305, USA*
[3] *Department of Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[4] *Department of Physics, University of Arizona, Tucson, AZ 85721, USA*
[5] *Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA*

ABSTRACT

The distribution of massive clusters of galaxies depends strongly on the total cosmic mass density, the mass variance, and the dark energy equation of state. As such, measures of galaxy clusters can provide constraints on these parameters and even test models of gravity, but only if observables of clusters can lead to accurate estimates of their total masses. Here, we carry out a study to investigate the ability of a blind spectroscopic survey to recover accurate galaxy cluster masses through their velocity dispersion. We focus on the Hobby Eberly Telescope Dark Energy Experiment (HETDEX), which will observe many hundreds of square degrees. HETDEX will employ the new VIRUS instrument, an integral field unit (IFU), covering the blue/optical portion of the spectrum (3500–5500 Å) blindly in two fields for a total of 420 degree^2 on the sky with fibers with a 1/4.5 fill factor. This allows surveys with VIRUS to measure redshifts for a large sample of galaxies out to $z < 0.5$ based on their absorption features or [O II] $\lambda3727$ emission (and Lyman-α over $1.9 < z < 3.5$). We use a detailed mock galaxy catalog from a semi-analytic model to simulate surveys observed with VIRUS, including: (1) a blind, HETDEX-like survey with an incomplete but uniform spectroscopic selection function; and (2) IFU surveys that target clusters directly, obtaining spectra of all galaxies in a VIRUS-sized field. For both surveys, we include realistic uncertainties from galaxy magnitude and line-flux limits. We benchmark both surveys against spectroscopic observations with "perfect" knowledge of galaxy line-of-sight velocities. Using both observing strategies, we produce cluster mass probability density functions $P(M_{200}|\vec{x})$ of a cluster mass (M_{200}) as a function of an observable parameter or set of observable parameters, $\vec{x}$, where we consider cases of where $\vec{x} = \{\sigma\}$, $= \{\sigma, z\}$, and $= \{\sigma, z, N_{gal}\}$ for velocity dispersion, σ, redshift, and number of observed galaxies, N_{gal}. This sentence will summarize how well HETDEX will do, and how that might be improved with targeted observations.

1 INTRODUCTION

Our ability to perform precision cosmology with clusters of galaxies has reached a critical turning point. The widely accepted ΛCDM 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, $P(M|z, X)$ the probability (P) that a galaxy cluster of given mass (M), located at redshift (z) determined using observable parameter (X). Generally, cluster mass calibrations fall into two distinct camps, simulation based and direct or statistical calibration.

One could use various simulations to attempt to calibrate this observable–mass relation (e.g., Vanderlinde et al. 2010; Sehgal et al. 2011). However, the primary challenge to this method is the incomplete understanding of the baryonic physics which take place in galaxy cluster environments. While there have been (and continue to be) many improve-

ments in the accuracy and power of simulations it is doubtful that in the coming years they will reach the accuracy level required where the observable–mass relation is dominated only by statistics (Weinberg et al. 2013).

The second broad camp is the direct calibration of cluster masses. This recipe has two distinct but not always independent tracks. The "direct" method uses direct observations of a small set of clusters and then uses known mass estimators, X-ray temperatures and luminosities (e.g., Mantz et al. 2010; Rykoff et al. 2014), cosmic microwave background (CMB) observations (e.g., Vanderlinde et al. 2010; Sehgal et al. 2011), optical richness (e.g., Abell 1958; Rykoff et al. 2012) or weak lensing (WL; e.g. Rozo et al. 2010) as examples, which provide a "true" mass. This directly calibrates the observable–mass relation which is then applied to a much larger sample. The complications lie in that the "true" masses are in fact estimations, and the methods used to recover these masses are subject to their own limitations. X-ray hydrostatic estimations assume hydrostatic equilibrium (e.g., Mantz et al. 2015) which may only be valid for a very small number and range of cluster masses. The Sunyaev–Zel'dovich Effect (SZE; Sunyaev & Zeldovich 1972), which uses the up–scattering of CMB photons to estimate cluster masses, provides accurate estimations of mass, but the ability to detect low mass galaxy clusters is currently limited by technology (e.g., Carlstrom et al. 2002). WL estimates are, in principle, correct in the mean, but they suffer from signal-to-noise requirements, limiting their usefulness in low mass clusters, and potentially suffer from line-of-sight effects as WL is sensitive to all mass along the line-of-sight. Virial mass estimators which determine the cluster mass based on the motions of the member galaxies (e.g., Ruel et al. 2014; Sifón et al. 2015b) are promising in that it is a direct measurement of the depth of clusters potential well, but suffers from systematics due to cluster formation physics which disrupts the velocity field.

The statistical method of determining galaxy cluster mass relies not on direct measurements of individual clusters but the calibration of observables for the entire sample which correlate with cluster mass. One example is the spatial clustering of the galaxy clusters themselves (e.g., Baxter et al. 2016). See Weinberg et al. (2013) for a comprehensive review. In practice, it will be a combination of the three methods touched on that will provide the most reliable determination of cluster masses.

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 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 $P(M|z, X)$. 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 (BAO) (Eisenstein et al. 2005) in the first moments of the universe, the survey will observe two fields for a total of 420 degree^2 (300 degree^2, Spring field and 120 degree^2, Fall field). Tuned to find Lyα 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 50 Virgo-sized ($M_{dyn} > 10^{15}\,\mathrm{M_\odot}$) clusters at $z < 0.5$. 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), DES, the *Spitzer*/HETDEX Exploratory Large-Area Survey (SHELA; Papovich et al. 2016), and the *Spitzer* IRAC Equatorial Survey (SpIES; Timlin et al. 2016, ApJ, submitted). 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 we settle on the verbage for everything else. 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: $\Omega_\Lambda = 0.714$, $\Omega_M = 0.286$, $\sigma_8 = 0.82$ and $H_0 = 70$ km s^{-1}Mpc^{-1}(taken from the Buzzard catalogs; see below), 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 Mock Catalogs

The Buzzard mock galaxy catalogs (R. Wechsler et al., private communication) cover 398.49 degree^2 between $4^h < RA < 6^h$ 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 `ROCK-STAR` halo finder (Behroozi et al. 2013) which also calculates halo masses.

Galaxy M_r luminosities are added to the velocity peaks using ShAM (Reddick et al. 2013), and ADDSEDs 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 mock catalog containing 238 million galaxies with $r < 29$ mag 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 galaxy, 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.

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 at central redshifts of 0.1, 0.2, and 0.4 using `HMFcalc` (Murray et al. 2013) and compare it 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. Could show a figure comparing the Buzzard catalog to a plot of the HMF for a couple of redshifts. Might be helpful, might not matter.

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, the right panel of Figure 1. 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. For the galaxies in the Buzzard catalog, 1.3% of the galaxies brighter than $g = 22$ mag (HETDEX's detection limit) have exactly zero [O II] luminosity. Of the 1.3%, only 3.3% (0.05% of all galaxies) have $g - r < 1.5$ suggesting the galaxies which are assigned zero [O II] luminosity are not significantly biasing our simulations.

The left 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 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. We could add a plot like the corner plot to show the multi-dimensionality of this process.
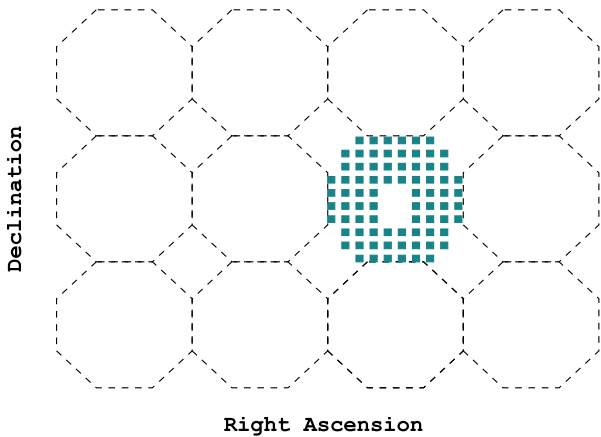
2.3 Mock Observations

HETDEX will perform blind spectroscopy (R ∼ 750 in 3500 − 5500 Å) over two fields along the celestial equator. The 300 degree^2, spring field and 120 degree^2 fall field will have no preselected targets. Using VIRUS on the 10-m Hobby-Eberly Telescope (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 reproduce the HETDEX IFU pattern, see Figure 2. The pattern consists of 78 IFUs, which are comprised of 448 optical fibers subtending a $50'' \times 50''$ region on the sky (Kelz et al. 2014). The inter-IFU spacing is also $50''$ spanning a total area of $16' \times 16'$ on the sky.
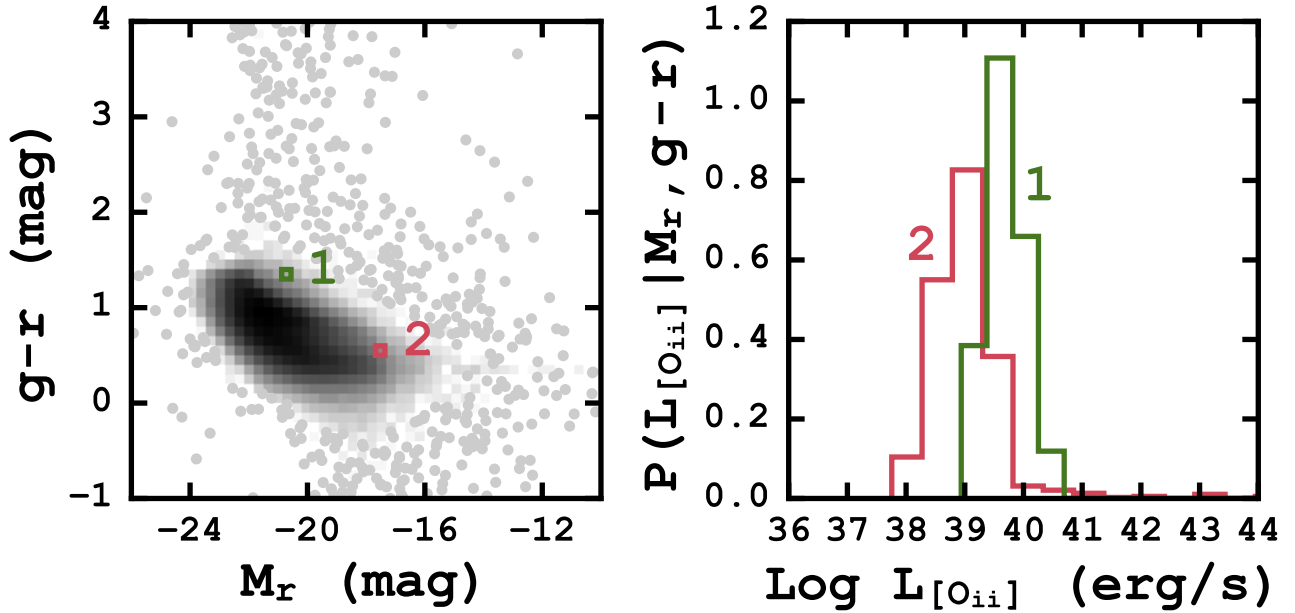
Figure 1. *Left*: CMD of 503113 $z < 0.2$ galaxies take from the SDSS DR12 where the shading scales with the density of points. The two colored boxes show regions containing potential catalog galaxies. *Right*: Probability histograms of the Log [O II] luminosity for the SDSS galaxies located in the two 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.

Figure 2. Representative observation tiling scheme for the HET-DEX $16' \times 16'$ 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.

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 398.49 degree^2 field of the Buzzard catalog we require 5370 pointings where 0.015 degree^2 of each pointing is covered by an IFU. The total area of the sky covered by an IFU is 80.80 degree^2 which gives a filling factor of 1/4.65 slightly decreased from the expected filling factor of 1/4.5.

The spectral coverage allows for the detection of [O II] ($\lambda\lambda 3727, 3729$) emitters to $z \sim 0.5$ and Ca H ($\lambda 3968.5$) and K ($\lambda 3933.7$) 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, targeted and survey-like. The targeted observations use "direct" observations where each cluster is targeted individually, 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. Both of these observations have HETDEX-like galaxy detection thresholds, for comparison we also include a set of targeted observations with "perfect" knowledge. Perfect observations assume no detection threshold, if a cluster member galaxy is observed, it is also detected. This provides an important best-case scenario, and differs from the true cluster properties because the recovered cluster properties are still calculated. These observations provide three levels of quality with "perfect knowledge" being the highest and survey 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 other 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 (LOSV) 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 (LOSV) to each galaxy, where

$$LOSV = c \frac{z - z_c}{1 + z_c} \tag{1}$$

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 (LOSVD) is calculated using a method of maximum likelihood following Walker et al. (2006). We maximize the probability function

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

where σ_p, $\langle \mu \rangle$, and σ_i is the LOSVD, the average radial velocity and the error on the individual LOSVs 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 using simple priors, $\langle \mu \rangle$ lies between the maximum and minimum LOSV and $0 < \sigma_p < 1400$ km s^{-1}. When the full distribution of LOSVDs is not used, the final LOSVD 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 LOSVD value bracketed by the error bars in

---

[1] http://dan.iel.fm/emcee/current/

$\sim 68\%$ 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 LOSV measurements, which will become important in future observations of real clusters (e.g., Boada 2016).

3.3 Dynamical Mass from Scaling Relations

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

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

with $A_{1D} = 1177 \pm 4.2$ 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). $H(z) = H_0 E(z)$ and $E(z) = \sqrt{\Omega_m(1 + z^3) + \Omega_\Lambda}$.

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 Other Estimates of Dynamical Mass

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 parameters and the target parameter (the cluster 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 $\vec{x} = \{\sigma, z, ...\}$ we make the joint probability be-
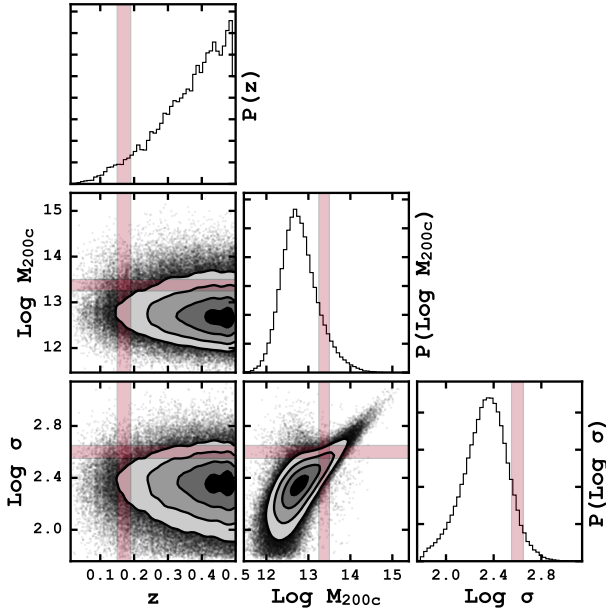
Figure 3. Corner plot of the *training* data with features σ and z. The corner plots shows all of the one and two dimensional posterior probability distributions used to determine the correct cluster mass. The colored rectangles, which have been enlarged for clarity, show the slices needed to create a conditional probability distribution of the mass, $P(M_{200c}|\vec{x})$. See text for a complete description.

tween the true cluster mass (M_{200c}) and $\vec{x}$. Figure 3, a corner plot, 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 $P(M_{200c}|\vec{x} = \{x_1, x_2, ...\})$ 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 $\sigma = 400$ 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 $P(M_{200c}|\vec{x} = \{\sigma = 400$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, $P(M_{200c})$, we combine the conditional probability distribution, $P(M_{200c}|\vec{x})$, created previously with the probability distribution of σ through Equation 4.

$$P(M_{200c}) = \int P(M_{200c}|\vec{x})P(\sigma)d\sigma \qquad (4)$$

The expected mass is determined by calculating the first moment of the probability density. This becomes our "predicted" mass, M_{pred}.

$$M_{pred} = \int M_{200c} P(M_{200c})dM_{200c} \qquad (5)$$

The confidence interval associated with this prediction can be estimated two ways. First, by calculating the second mo-

ment of the probability density through

$$\text{Var} = \int (M_{200c} - M_{pred})^2 P(M_{200c})dM_{200c} \qquad (6)$$

or by drawing many samples from $P(M_{200c})$ and calculating the values at the 16th and 84th percentile. In practice we find that both methods produce similar results for a large number 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 cluster mass 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 shorten to just random forest (RF). Decision trees can be visualized 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 uncertainties on the different moments of a population, a prediction interval is unique to each prediction. In many regresson analyses, such as linear fitting, the prediction intervals are based on underlying assumptions of normally 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, targeted, and survey. Targeted observations are direct observations of a cluster where each cluster member galaxy, above the detection
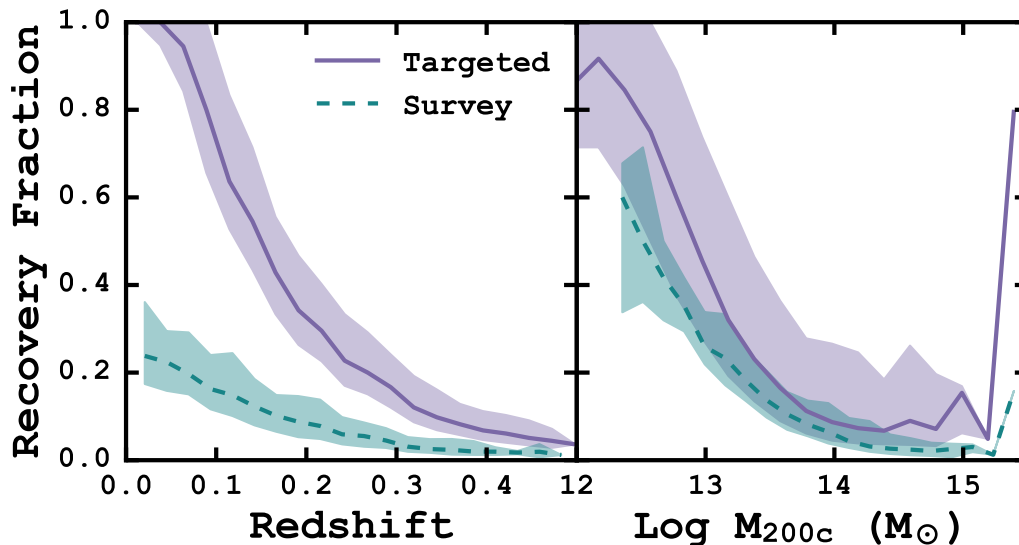
Figure 4. Recovery fractions (N_{obs}/N_{True}) of cluster member galaxies as a function of redshift and true cluster mass for the targeted and survey observing strategies. We have applied HETDEX-like observational limits on the cluster galaxy detection, and require at least five galaxies to be detected for a cluster to be recovered. 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.

thresholds (see Section 2.3), is observed. Survey observations mimic the HETDEX observation strategy such that no cluster is directly observed, and only the cluster member galaxies above the detection threshold and within an IFU (see Figure 2) are observed. 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 clusters 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 fewer than five member galaxies are observed ($N_{obs} < 5$) 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 (left panel of Figure 4), there are two effects at work. The decrease in recovery fraction with increasing redshift is a magnitude effect. We check this by limiting the cluster galaxy detection by absolute magnitude which increases the recovery fraction to > 70% at all redshifts, implying the decline is a result of the apparent magnitude cluster galaxy detection threshold. 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 the low mass clusters we detect ($N_{obs} > 5$), we 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, as above. Also, high mass clusters are rare, so in order to detect them we must probe a large volume of space. The higher redshift cluster members suffer from the limiting apparent magnitude and suppress the recovery fraction at fixed mass. If we were to limit the survey to $z < 0.2$ we find the recovery fraction of clusters, at all masses, increases substantially, and we find a much more consistent detection fraction across all masses.

4.2 Mass estimates

In this section we discuss the how accurately we recover the true cluster mass from a set of observations. We report on three methods, the power law based approach (Eq. 3), the probability based approach (Section 3.4.1) and the ML based method (Section 3.4.2). For each method we consider observations with perfect knowledge, targeted observations and survey observations.

Because it represents the best possible scenario, the perfect knowledge observations 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 average bias (given in Table 1)

$$\mu_{\text{bias}}(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} (\hat{y_i} - y_i). \tag{7}$$

where y are the true values and $\hat{y}$ are the predicted values,
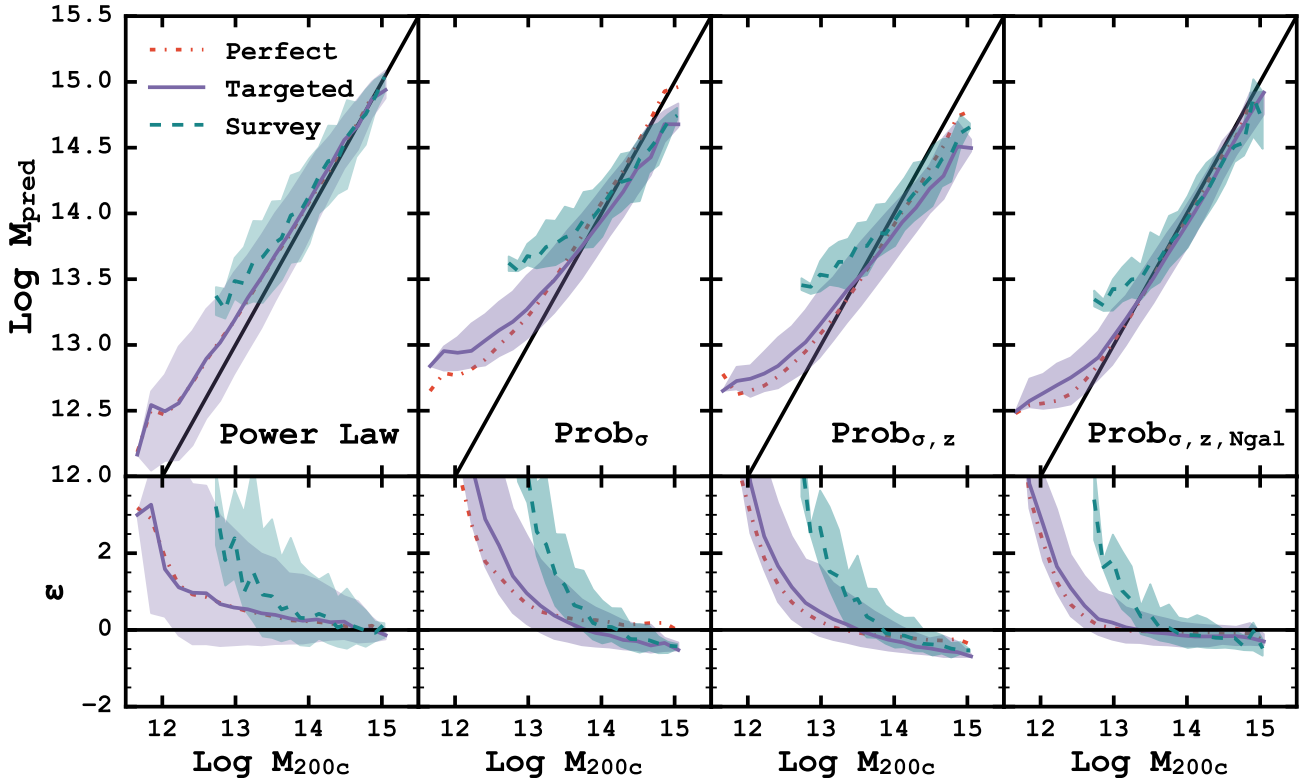
Figure 5. Mass predictions for the power law scaling relation (Equation 3) and the probability based technique with different input features as a function of true cluster mass. The bottom row of panels shows the fractional error (Equation 9) also as a function of true cluster mass. The solid black line 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.

and the scatter about the bias (given in Table 2)

$$\sigma_{bias}(y, \hat{y}, \mu_{bias}) = \left[\frac{1}{N-1} \sum_{i=0}^{N} (\hat{y}_i - y_i - \mu_{bias})^2 \right]^{1/2} \quad (8)$$

for different bins of cluster mass. Both metrics evaluate how closely the ensemble of predicted cluster masses are to the true cluster masses, and in both cases values closer to zero are better.

We begin with the perfect knowledge observations. These observations are of the same clusters as the targeted observations but without any observational limits. The cluster masses predicted by Equation 3 gives the following results. Across all cluster masses, we find $\mu_{bias} = 0.15 \pm 0.002$ dex. The scatter in recovered masses can be attributed to both physical and numerical effects. The presence of any in-falling matter onto lower mass clusters can introduce a significant amount of substructure, which increases the observed LOSVD increasing the predicted mass (e.g., Ntampaka et al. 2015). 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 cluster 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 pre-

sented in this section represent an upper bound (the best) on the accuracy achievable through this method.

For the targeted and survey observations the power law predicted cluster masses give $\mu_{bias} = 0.13 \pm 0.003$ dex and $\mu_{bias} = 0.11 \pm 0.009$ dex, respectively. So for the clusters that we detect with survey observations, we obtain similar levels of accuracy as to the targeted observations, on the average. This does not mean that the survey observations cannot be improved by targeted observations. In fact, when comparing only the galaxies which overlap between the two samples the bias and scatter of the targeted observations is significantly decreased as more cluster member galaxies are detected, better sampling the LOSVD PDF. The targeted observations performs similarly on the average because many lower mass clusters are included in the sample, increasing the bias of the overall sample.

In both Figures 5 and 6, we show the predicted (M_{pred}) versus true (M_{200c}) cluster masses for each of the two observing strategies. In each panel the solid black line is the 1:1 relationship, the solid purple line is the median recovered mass for the targeted observing, and the turquoise, 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). For reference we also show the median masses recovered with perfect
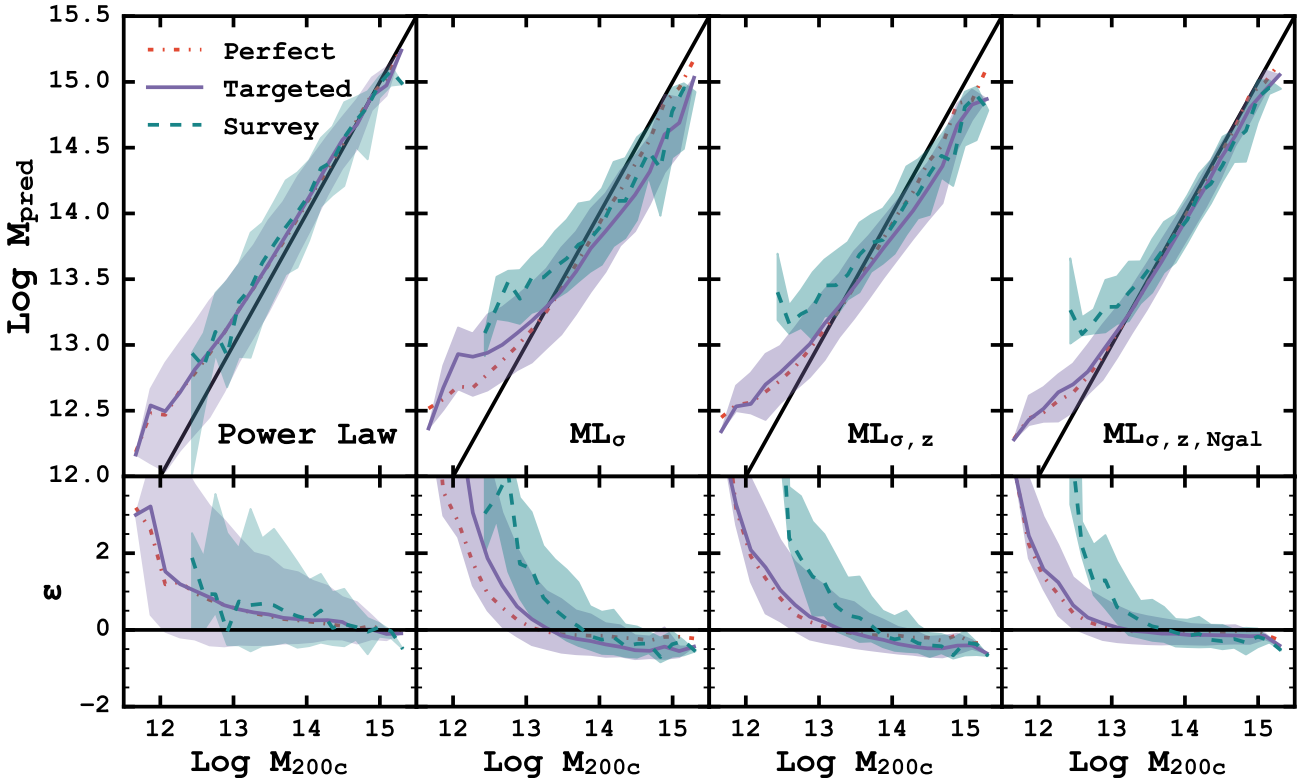
Figure 6. Mass predictions for the power law scaling relation (Equation 3) and the ML based technique with different input features as a function of true cluster mass. The bottom row of panels shows the fractional error (Equation 9) also as a function of true cluster mass. The solid black line shows 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.

knowledge as a solid orange line. The lower panels show the fractional cluster mass error:

$$\epsilon = (M_{pred} - M_{200c})/M_{200c} \qquad (9)$$

where M_{pred} is the predicted cluster mass and M_{200c} is the true cluster mass.

Qualitatively, the top panels of Figures 5 and 6 show that both the probability based and ML based methods out perform (closer to the black 1:1 relation) the power law method when taking advantage of other cluster observables (z, N_{gal}, etc.). Generally, we find that the single parameter probability and ML methods perform significantly poorer than the power law method, especially at low cluster masses. When combined with the cluster redshift, the predicted cluster masses are improved, even though it is unclear exactly why the additional information should effect the predicted cluster mass. The final addition of the number of observed galaxies, N_{gal} acts as a type of richness estimate, and significantly improves both the bias and the amount of scatter in the predicted masses.

We quantify the bias and scatter for all of the different cluster mass recovery strategies and observing methods in Table 1 and Table 2. It should serve as a type of look up table for future cluster observations with HETDEX. The columns represent bins of predicted galaxy cluster mass and the individual values show the bias and scatter of the true

cluster mass. The three horizontal sections represent perfect, targeted and survey observations respectively. So, for example, if a cluster mass is predicted using the $ML_{\sigma,z}$ method and targeted observations to be Log M = 13 − 13.5 $M_\odot$, it is biased upward by 0.03 ± 0.00 dex.

A few caveats apply to the numbers given in Table 1 and Table 2. While we provide corrections for cluster masses above 10^{15} $M_\odot$, they are estimated from only a handful of objects, and do not constitute a representative sample of clusters. On the oposite end of the cluster mass spectrum, survey observations of low mass clusters are incomplete for two potential reasons. There are very few, if any, clusters detected with survey observations below 10^{12} $M_\odot$. Missing corrections above 10^{12} $M_\odot$ in the probability based methods are due to low number statstics and an inability to sample the cluster mass PDF well enough to assign cluster mass predictions.

4.2.1 Significance of Cosmology Choice

The underlying simulation used to produce the Buzzard catalogs, as with every simulation, makes specific assumptions about the cosmology used and how the galaxies are populated into the dark matter halos. When using simulation data to train ML methods, we incorporate all of those as-

Table 1. Mean bias (M_{pred} - M_{200c}) for different bins of predicted cluster mass. This table shows the bias in the predicted cluster mass for the perfect (top section), targeted (middle section), and survey (bottom section) observations in different predicted mass bins. The different mass recovery strategies are given in the leftmost column. It can be used to understand how the predicted cluster mass differs from the true cluster masses. Positive numbers indicate the predicted cluster mass over estimates when compared to the true cluster mass.

	Method	Bins – Log M_{pred}							
		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	0.49 ± 0.09	0.28 ± 0.02	0.23 ± 0.01	0.16 ± 0.00	0.11 ± 0.00	0.07 ± 0.00	0.02 ± 0.01	−0.07 ± 0.04
	σ	0.87 ± 0.04	0.53 ± 0.01	0.32 ± 0.01	0.16 ± 0.00	0.10 ± 0.00	0.07 ± 0.00	0.05 ± 0.01	−0.18 ± 0.06
	σ, z	0.76 ± 0.04	0.42 ± 0.01	0.17 ± 0.00	0.01 ± 0.00	−0.06 ± 0.00	−0.11 ± 0.01	−0.14 ± 0.01	−0.38 ± 0.13
	σ, z, N_{gal}	0.65 ± 0.03	0.28 ± 0.01	0.04 ± 0.01	−0.02 ± 0.00	−0.02 ± 0.00	−0.05 ± 0.01	−0.22 ± 0.12	−6.62 ± 3.06
ML Based	Power Law	0.49 ± 0.09	0.28 ± 0.02	0.23 ± 0.01	0.16 ± 0.00	0.11 ± 0.00	0.07 ± 0.00	0.02 ± 0.01	−0.07 ± 0.04
	σ	0.69 ± 0.05	0.36 ± 0.01	0.15 ± 0.01	−0.02 ± 0.00	−0.07 ± 0.00	−0.09 ± 0.01	−0.11 ± 0.01	−0.12 ± 0.07
	σ, z	0.62 ± 0.05	0.33 ± 0.01	0.12 ± 0.00	−0.01 ± 0.00	−0.06 ± 0.00	−0.09 ± 0.00	−0.11 ± 0.01	−0.19 ± 0.05
	σ, z, N_{gal}	0.51 ± 0.04	0.23 ± 0.01	0.05 ± 0.00	−0.02 ± 0.00	−0.02 ± 0.00	−0.02 ± 0.00	−0.02 ± 0.01	−0.08 ± 0.03
Prob Based	Power Law	0.49 ± 0.08	0.27 ± 0.02	0.20 ± 0.01	0.13 ± 0.01	0.10 ± 0.01	0.09 ± 0.01	0.02 ± 0.01	−0.08 ± 0.03
	σ	1.02 ± 0.03	0.66 ± 0.01	0.40 ± 0.00	0.17 ± 0.00	0.02 ± 0.00	−0.08 ± 0.01	−0.19 ± 0.01	−0.35 ± 0.10
	σ, z	0.85 ± 0.05	0.49 ± 0.01	0.25 ± 0.00	0.08 ± 0.00	−0.06 ± 0.00	−0.21 ± 0.02	−0.35 ± 0.02	−0.59 ± 0.12
	σ, z, N_{gal}	0.70 ± 0.03	0.37 ± 0.01	0.13 ± 0.00	0.01 ± 0.00	−0.05 ± 0.00	−0.12 ± 0.02	−0.44 ± 0.18	−4.45 ± 2.81
ML Based	Power Law	0.49 ± 0.08	0.27 ± 0.02	0.20 ± 0.01	0.13 ± 0.01	0.10 ± 0.01	0.09 ± 0.01	0.02 ± 0.01	−0.08 ± 0.03
	σ	0.76 ± 0.07	0.54 ± 0.01	0.26 ± 0.01	0.01 ± 0.00	−0.13 ± 0.00	−0.23 ± 0.01	−0.33 ± 0.02	−0.32 ± 0.03
	σ, z	0.66 ± 0.03	0.37 ± 0.01	0.17 ± 0.01	0.03 ± 0.00	−0.09 ± 0.00	−0.21 ± 0.01	−0.31 ± 0.02	−0.28 ± 0.05
	σ, z, N_{gal}	0.57 ± 0.03	0.28 ± 0.01	0.09 ± 0.00	−0.01 ± 0.00	−0.05 ± 0.00	−0.08 ± 0.00	−0.07 ± 0.01	−0.14 ± 0.05
Prob Based	Power Law	...	...	0.53 ± 0.04	0.30 ± 0.02	0.16 ± 0.01	0.07 ± 0.01	0.01 ± 0.02	−0.09 ± 0.05
	σ	...	...	0.77 ± 0.03	0.42 ± 0.01	0.18 ± 0.01	−0.03 ± 0.01	−0.18 ± 0.02	−0.39 ± 0.08
	σ, z	...	...	0.61 ± 0.03	0.29 ± 0.01	0.08 ± 0.01	−0.11 ± 0.01	−0.38 ± 0.12	−0.48 ± 0.10
	σ, z, N_{gal}	...	...	0.48 ± 0.04	0.18 ± 0.01	0.02 ± 0.01	−0.08 ± 0.01	−0.50 ± 0.20	−8.81 ± 3.02
ML Based	Power Law	...	0.06 ± 0.46	0.17 ± 0.07	0.17 ± 0.02	0.13 ± 0.01	0.07 ± 0.01	0.01 ± 0.02	−0.09 ± 0.05
	σ	...	0.58 ± 0.10	0.54 ± 0.04	0.24 ± 0.02	0.03 ± 0.01	−0.17 ± 0.01	−0.26 ± 0.03	−0.30 ± 0.10
	σ, z	...	1.02 ± 0.20	0.42 ± 0.04	0.19 ± 0.01	0.03 ± 0.01	−0.14 ± 0.01	−0.25 ± 0.02	−0.31 ± 0.10
	σ, z, N_{gal}	...	0.91 ± 0.26	0.39 ± 0.03	0.12 ± 0.01	−0.00 ± 0.01	−0.08 ± 0.01	−0.14 ± 0.02	−0.21 ± 0.08

sumptions into the learned feature associations. This could lead to a notable bias when applying the ML method to data (real or simulated) which has significantly different cosmological assumptions.

To investigate this, we select cluster halo and galaxy catalogs from the Millennium simulation (Springel et al. 2005) obtained via querying the Millennium online database[2] (Lemson & The Virgo Consortium 2006). The Millennium simulation tracks 2160^3 dark matter particles of 8.6×10^8 h^{-1} $M_\odot$ inside a comoving 500 $(h^{-1}\mathrm{Mpc})^3$ box from $z = 127$ to 0. It adopts a flat cosmological model based of the values derived from the Two-degree Field Galaxy Redshift Survey (Colless et al. 2001) and the first year data of the *Wilkinson Microwave Anisotropy Probe* (WMAP; Spergel et al. 2003): $\Omega_\Lambda = 0.75$, $\Omega_M = 0.25$, $\sigma_8 = 0.9$, $n_s = 1$ and $H_0 = 73$ km s^{-1}Mpc^{-1}.

The halo catalogs are created using a friends-of-friends algorithm (e.g., Turner & Gott, J. R. 1976) with linking length, $b = 0.2$. The halos are populated with galaxies using a semi-analytical model described in Croton et al. (2006); De Lucia et al. (2006); De Lucia & Blaizot (2007). We select 4806 clusters, comprised of 623663 galaxies, at $0.02 < z < 0.5$ and $M > 10^{13}$ $M_\odot$, and apply the same data processing as with the Buzzard galaxies. We begin by

assigning each galaxy an [O II] flux value (see Section 2.2), and "observe" each galaxy using realistic (see Section 2.3) observational limits. After recovering 3750 clusters which have at least five galaxies observed, we calculate the LOSVD of each cluster as in Section 3.2.

Cluster masses are assigned using the ML methods (see Section 3.4.2) using two different training sets. We use the full set of clusters detected in the Buzzard catalogs (14000 clusters with $M > 10^{11}$ $M_\odot$) and a training-testing split of the Millennium clusters. The top pandel of Figure 7 shows the ML predicted cluster masses as a function of true cluster mass for the ~4000 Millennium clusters with the two different training sets. The median (solid and dashed lines) predicted cluster masses show similar trends regardless of the training data set used. Figure 7's bottom panel show the fraction error (Equation 9) also as a function of true cluster mass. The large amount of scatter (the shaded area) in the fractional error for the Buzzard trained predictions is due to the training set including clusters with masses below the $M = 10^{13}$ $M_\odot$ threshold for the Millennium clusters. This allows the ML method to predict masses which can be significantly different, whereas the Millennium training set does not include $M < 10^{13}$ $M_\odot$ clusters, which reduces the scatter of the predicted masses.

Maybe a sentence or two that specifically say we don't think this is a problem.

---

[2] http://gavo.mpa-garching.mpg.de/Millennium/
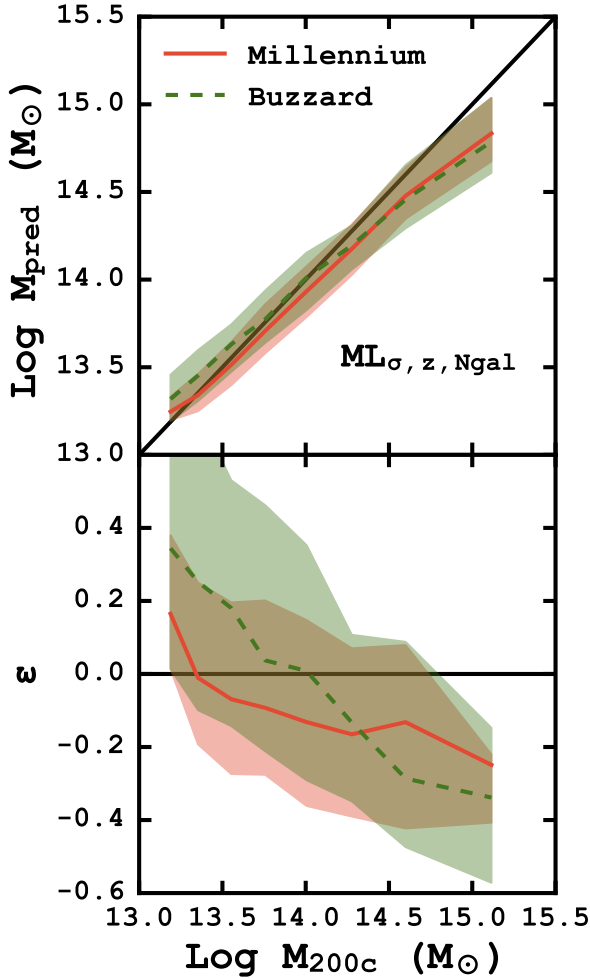
Figure 7. *Top:* ML based cluster mass predictions for the Millennium simulation clusters where the ML method has been trained with either a subset of the Millennium clusters (solid line) or the Buzzard catalog (dashed line). The shaded areas show the 68% scatter around the median. The solid black line shows the 1:1 relation.The solid black line shows the 1:1 relation. *Bottom:* The fractional error (Equation 9) also as a function of true cluster mass. The similarity of the predictions with the different training sets demonstrates how the ML method is not sensitive to the underlying cosmological assumptions.

5   HETDEX AS A GALAXY CLUSTER SURVEY AT $Z < 0.5$

5.1   Impact on Cluster Cosmology

Galaxy clusters are sensitive probes of Ω_m, the total mass $(\Omega_b + \Omega_c)$ density in the universe, by tracing the peaks in the universal matter density often referred to as the power spectrum of matter density fluctuations or the matter power spectrum and $\sigma_8 h^{-1}$, the normalization of the power spectrum, by the comparison of the number density of observed halos to that predicted in cosmological models. Although, in reality, one measures $\sigma_8 h^{-1} \Omega_m^q$, where the value of q depends on the masses and redshifts of the halos considered.

To get a sense of how well HETDEX will be able to

constrain cosmological parameters we follow the discussion of Weinberg et al. (2013), and begin with a few simplifying assumptions. While sensitive to Ω_m, the number density of clusters do not necessarily provide the strongest contraint, so we assume that some combination of other data sets (e.g., CMB, BAO, supernovae, WL, etc.) will constrain Ω_m to higher precision than is possible with clusters alone.

To estimate the error associated with a measurement of $\sigma_8 h^{-1}$ (which Weinberg et al. 2013 refer to as $\sigma_{11,abs}$), Weinberg et al. (2013) consider two sources of uncertainty, the systematic uncertainties in cluster mass calibration and the statistical uncertainty in the observed number density of clusters. The authors combine these two uncertainties though (their Eq. 141):

$$\Delta \ln \sigma_8 h^{-1}(z) \approx q(z) \times \max \left[\Delta \ln M, \alpha(z)^{-1} \Delta \ln N \right]. \quad (10)$$

where q is the degeneracy exponent between $\sigma_8 h^{-1}$ and Ω_m, $\Delta \ln M$ is the mass scale uncertainty, $\Delta \ln N$ is the cluster statistical uncertainty, and α is slope of the cumulative HMF. Using the Tinker et al. (2008) HMF at $z \sim 0.2$ and a limiting cluster mass of 10^{14} $M_\odot$, Weinberg et al. (2013) estimate $q \sim 0.4$, $\alpha \sim 3$, and find that any cluster survey with more than 10-20 clusters is dominated by the uncertainty in the overall mass scale.

For a survey such as HETDEX, we can estimate the constraints on $\sigma_8 h^{-1}$ using Equation 10. Considering clusters with masses above 10^{14} $M_\odot$, with perfect knowledge, and the ML based cluster mass estimates the mass scale uncertainty is reduced to ~ 0.075 dex or about 20%. This gives a uncertainty on $\sigma_8 h^{-1}$ of 7%. For clusters above 10^{14} $M_\odot$, survey observations can constrain the masses to about 51% which, in turn, constrains $\sigma_8 h^{-1}$ to 20%.

Because of the simplifying assumptions, and the superior quality of the data (no contamination, signal-to-noise issues, etc.), realistic expectations for HETDEX is to directly constrain $\sigma_8 h^{-1}$ to better precision than other methods (e.g., CMB, WL, X-ray) seems bleak. For the ~ 1500 clusters detected with survey observations to constrain $\sigma_8 h^{-1}$ and to be dominated by cluster statistics alone ($\Delta \ln N \sim N^{-1/2}$), the absolute cluster calibration would need to be better than 2.5%. For a fully targeted survey, about 14000 clusters, this cluster mass calibration uncertainty reduces to $> 1\%$. Here I want to say a word or two about how all hope isn't lost. We can still use HETDEX to reduce the scatter on the richness-mass relationship which can help other estimators get down to the levels we need to really nail down these numbers.

5.2   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

Table 2. Scatter in cluster mass after bias correction for different bins of predicted cluster mass. This table shows the scatter in the predicted cluster mass for the perfect (top section), targeted (middle section), and survey (bottom section) observations in different predicted mass bins. The different mass recovery strategies are given in the leftmost column. It can be used to understand how the predicted cluster mass differs from the true cluster masses.

	Method	Bins – Log M_{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	0.36 ± 0.07	0.39 ± 0.01	0.34 ± 0.00	0.22 ± 0.00	0.16 ± 0.00	0.14 ± 0.00	0.14 ± 0.01	0.11 ± 0.04
	σ	0.16 ± 0.03	0.24 ± 0.01	0.26 ± 0.00	0.20 ± 0.00	0.16 ± 0.00	0.14 ± 0.00	0.16 ± 0.01	0.19 ± 0.07
	σ, z	0.17 ± 0.03	0.21 ± 0.01	0.23 ± 0.00	0.19 ± 0.00	0.16 ± 0.00	0.55 ± 0.01	0.16 ± 0.01	0.39 ± 0.14
	σ, z, N_{gal}	0.10 ± 0.02	0.14 ± 0.01	0.47 ± 0.01	0.14 ± 0.00	0.10 ± 0.00	0.54 ± 0.01	1.62 ± 0.09	9.31 ± 3.38
ML Based	Power Law	0.36 ± 0.07	0.39 ± 0.01	0.34 ± 0.00	0.22 ± 0.00	0.16 ± 0.00	0.14 ± 0.00	0.14 ± 0.01	0.11 ± 0.04
	σ	0.22 ± 0.04	0.27 ± 0.01	0.29 ± 0.00	0.24 ± 0.00	0.21 ± 0.00	0.19 ± 0.00	0.18 ± 0.01	0.22 ± 0.08
	σ, z	0.19 ± 0.04	0.22 ± 0.01	0.25 ± 0.00	0.20 ± 0.00	0.18 ± 0.00	0.16 ± 0.00	0.17 ± 0.01	0.15 ± 0.05
	σ, z, N_{gal}	0.15 ± 0.03	0.18 ± 0.01	0.17 ± 0.00	0.14 ± 0.00	0.10 ± 0.00	0.08 ± 0.00	0.08 ± 0.00	0.10 ± 0.03
Prob Based	Power Law	0.33 ± 0.07	0.41 ± 0.01	0.43 ± 0.01	0.39 ± 0.00	0.33 ± 0.00	0.27 ± 0.01	0.18 ± 0.01	0.11 ± 0.04
	σ	0.14 ± 0.03	0.19 ± 0.01	0.24 ± 0.00	0.25 ± 0.00	0.25 ± 0.00	0.22 ± 0.00	0.19 ± 0.01	0.30 ± 0.11
	σ, z	0.19 ± 0.04	0.20 ± 0.01	0.24 ± 0.00	0.24 ± 0.00	0.23 ± 0.00	0.56 ± 0.01	0.21 ± 0.01	0.36 ± 0.13
	σ, z, N_{gal}	0.14 ± 0.03	0.18 ± 0.01	0.20 ± 0.00	0.19 ± 0.00	0.17 ± 0.00	0.67 ± 0.01	2.28 ± 0.13	8.56 ± 3.11
ML Based	Power Law	0.33 ± 0.07	0.41 ± 0.01	0.43 ± 0.01	0.39 ± 0.00	0.33 ± 0.00	0.27 ± 0.01	0.18 ± 0.01	0.11 ± 0.04
	σ	0.29 ± 0.06	0.29 ± 0.01	0.31 ± 0.00	0.31 ± 0.00	0.32 ± 0.00	0.30 ± 0.01	0.29 ± 0.02	0.11 ± 0.04
	σ, z	0.14 ± 0.03	0.23 ± 0.01	0.26 ± 0.00	0.25 ± 0.00	0.24 ± 0.00	0.22 ± 0.00	0.25 ± 0.01	0.15 ± 0.05
	σ, z, N_{gal}	0.13 ± 0.03	0.21 ± 0.01	0.21 ± 0.00	0.18 ± 0.00	0.16 ± 0.00	0.14 ± 0.00	0.12 ± 0.01	0.15 ± 0.05
Prob Based	Power Law	...	...	0.16 ± 0.03	0.29 ± 0.01	0.33 ± 0.01	0.31 ± 0.01	0.25 ± 0.02	0.15 ± 0.06
	σ	...	...	0.11 ± 0.02	0.18 ± 0.01	0.22 ± 0.01	0.22 ± 0.01	0.19 ± 0.01	0.25 ± 0.09
	σ, z	...	...	0.13 ± 0.03	0.19 ± 0.01	0.22 ± 0.01	0.21 ± 0.01	1.31 ± 0.08	0.32 ± 0.11
	σ, z, N_{gal}	...	...	0.14 ± 0.03	0.18 ± 0.01	0.19 ± 0.01	0.19 ± 0.01	2.27 ± 0.15	9.19 ± 3.34
ML Based	Power Law	...	1.41 ± inf	0.40 ± 0.05	0.41 ± 0.02	0.38 ± 0.01	0.32 ± 0.01	0.25 ± 0.02	0.15 ± 0.06
	σ	...	0.30 ± inf	0.27 ± 0.03	0.28 ± 0.01	0.30 ± 0.01	0.30 ± 0.01	0.30 ± 0.02	0.29 ± 0.11
	σ, z	...	0.61 ± inf	0.22 ± 0.03	0.25 ± 0.01	0.25 ± 0.01	0.25 ± 0.01	0.25 ± 0.02	0.30 ± 0.11
	σ, z, N_{gal}	...	0.80 ± inf	0.19 ± 0.02	0.21 ± 0.01	0.20 ± 0.01	0.19 ± 0.01	0.19 ± 0.01	0.23 ± 0.08

fixed optical richness are imprecisely known (Rykoff et al. 2012). These systematics remain the major source of uncertainty 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 HET-DEX 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 8 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

$$\ln L = \frac{-1}{2}\left(\sum_{i=1}^{N}\frac{[y_i - mx_i - b]^2}{\sigma_{y,i}} + \sum_{i=1}^{N}\frac{[y_i - mx_i - b]^2}{\sigma_{x,i}}\right) \quad (11)$$

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

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

and the survey observations as

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

The bottom panel of Figure 8 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
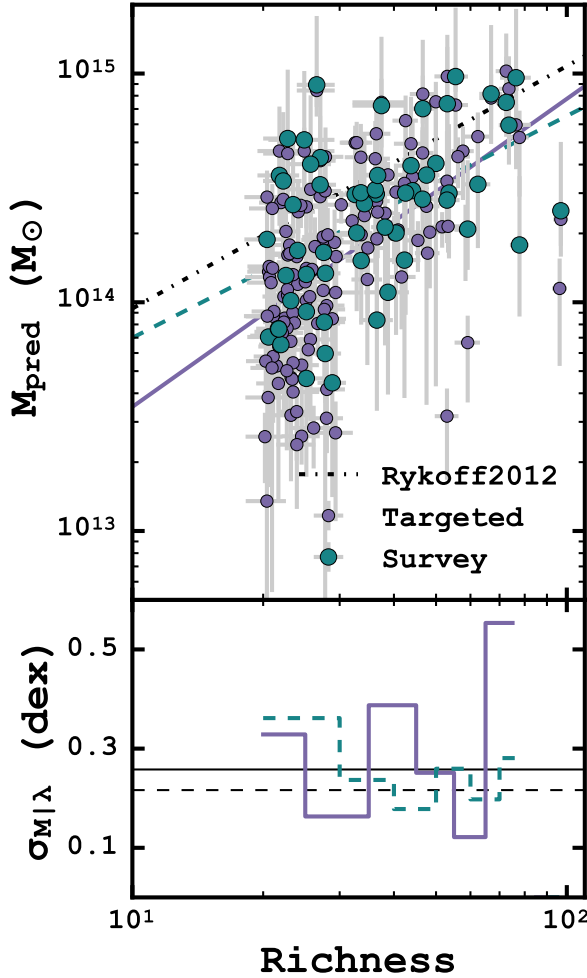
Figure 8. The optical richness, λ, versus the predicted cluster mass. The purple and blue points represent clusters with targeted and survey detections respectively. Error bars show 1σ prediction interval. The solid and dashed lines are fits to either the targeted or survey sets.

the targeted to survey observations is due to fewer objects detected in survey mode.

6 SUMMARY

Here, we present detailed simulations of the upcoming HET-DEX 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 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 $13 < LogM < 14.5$ $M_\odot$, we find the $ML_{\sigma,z,N_{gal}}$ method performs the best. Below 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 Log M < 13 $M_\odot$ from survey observations should be targeted. Galaxy clusters with inferred masses below 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 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 and reproducible science, 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

REFERENCES

Abell G. O., 1958, ApJS, 3, 211

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

Baxter E. J., Rozo E., Jain B., Rykoff E., Wechsler R. H., 2016, eprint arXiv:1604.00048, p. 7

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

Boada S., 2016, MNRAS

Bocquet S., et al., 2015, ApJ, 799, 214

Carlstrom J. E., Holder G. P., Reese E. D., 2002, Annu. Rev. Astron. Astrophys., 40, 643

Carlstrom J. E., et al., 2011, PASP, 123, 568

Chabrier G., 2003, PASP, 115, 763

Colless M., et al., 2001, MNRAS, 328, 1039

Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369

Croton D. J., et al., 2006, MNRAS, 365, 11

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, MNRAS, 366, 499

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

Lemson G., The Virgo Consortium 2006, eprint arXiv:astro-ph/0608019

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=10.1.1.48.886{&}rep=rep1{&}type=pdf. Department of Statistics, University of Toronto, Toronto, doi:10.1.1.48.886, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.886{&}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

Papovich C., et al., 2016, eprint arXiv:1603.05660, p. 17

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...34R

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

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

Spergel D. N., et al., 2003, ApJS, 148, 175

Springel V., 2005, MNRAS, 364, 1105

Springel V., et al., 2005, Nature, 435, 629

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

Turner E. L., Gott, J. R. I., 1976, ApJS, 32, 409

Vanderlinde K., et al., 2010, ApJ, 722, 1180

Walker M. G., Mateo M., Olszewski E. W., Bernstein R., Wang X., Woodroofe M., 2006, AJ, 131, 2114

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87

van der Burg R. F. J., Muzzin A., Hoekstra H., Wilson G., Lidman C., Yee H. K. C., 2014, A&A, 561, A79