

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

Steven Boada,<sup>1\*</sup> C. Papovich,<sup>1</sup> R. Wechsler<sup>2,3</sup>

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

<sup>2</sup>*Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, Stanford, CA 94305, USA*

<sup>3</sup>*Department of Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

ABSTRACT GOES HERE!! Lorem ipsum dolor sit amet, consectetur adipisicing 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.

## 1 INTRODUCTION

Our ability to perform precision cosmology with clusters of galaxies has reached a critical turning point. The widely accepted  $\Lambda$ 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 observations (e.g., [Vanderlinde et al. 2010](#); [Sehgal et al. 2011](#)) using the Sunyaev–Zeldovich 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 degree<sup>2</sup> from two fields (300 degree<sup>2</sup>, Spring field and 120 degree<sup>2</sup>, Fall field). Tuned to find Ly $\alpha$  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 ( $M_{dyn} \sim 10^{15} M_{\odot}$ ) 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.

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.

#### **Give outline of paper section.**

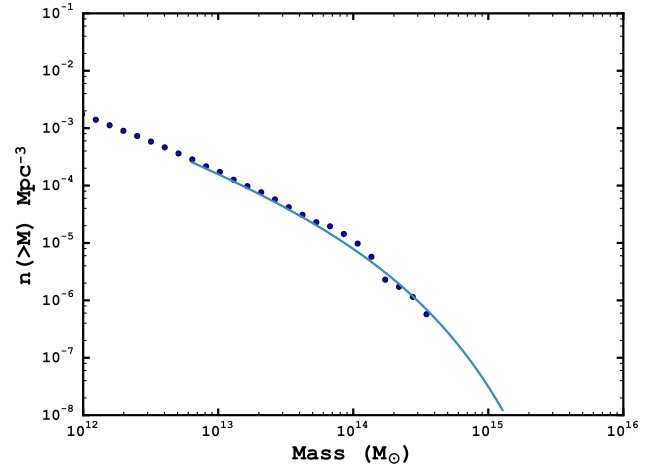
Throughout this paper, we adopt the following cosmological model ( $\Omega_\Lambda = 0.77$ ,  $\Omega_M = 0.23$ ,  $\sigma_8 = 0.83$  and  $H_0 = 72 \text{ km s}^{-1}\text{Mpc}^{-1}$ ), assume a Chabrier initial mass function (IMF; Chabrier 2003), and use AB magnitudes (Oke 1974).

## 2 DATA AND MOCK OBSERVATIONS

Blah blah intro stuff...

### 2.1 The “Buzzard” Catalogs

The “Buzzard” mock galaxy catalogs (R. Wechsler et al., private communication) cover  $375.68 \text{ degree}^2$  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.



**Figure 1.** The cumulative MF of halos above  $M_{200c}$  at  $z = 0.1$ . The predicted MF is from tinker2008.

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  $398.49 \text{ sq. degree}$  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  $z \leq 8.7$ .

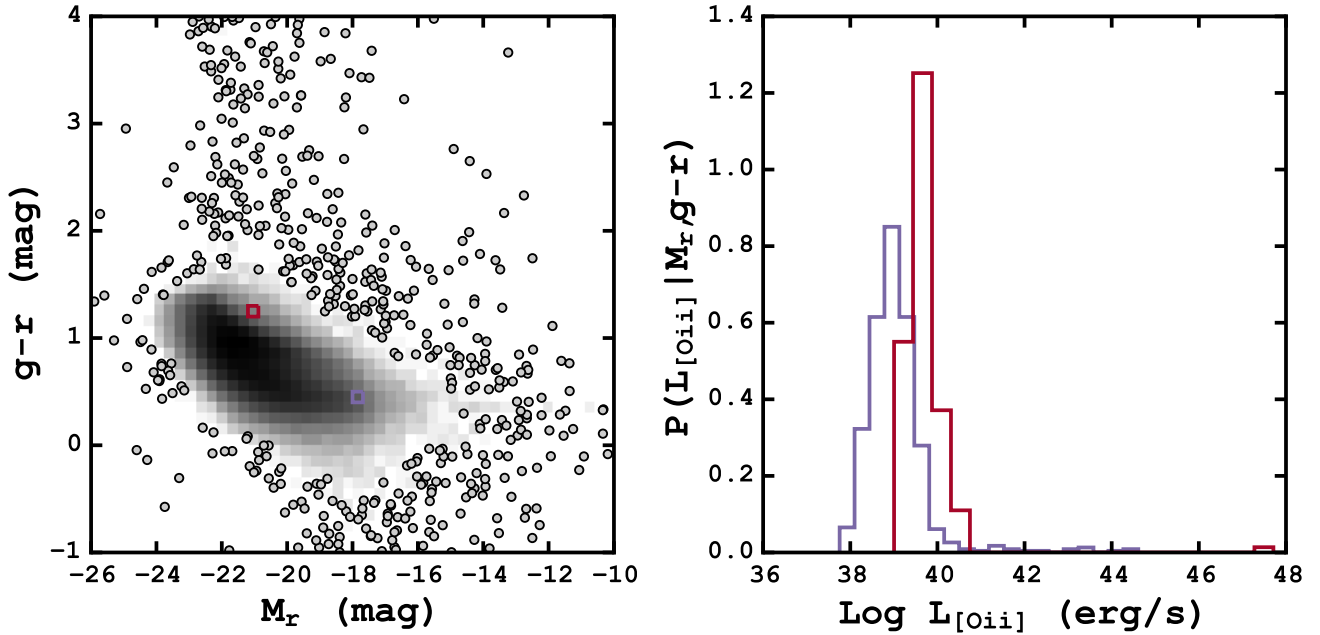
The catalog information, used in this study, is broken into two large portions. The “truth” files contain the characteristics of each individual galaxies, such as right ascension (RA), declination (DEC), redshift ( $z$ ), observed and rest-frame magnitudes, and many others. The “halo” files contain information 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 on emission or absorption lines or estimations of whether the halo is relaxed or not. We supplement the catalogs with this information and describe the method in Section 2.2 and others.

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). Shown in Figure 1 the HMF is calculated at central redshifts of 0, 0.2, and 0.4 using HMFcalc (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 expected HMF and the observed.

### 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



**Figure 2.** *Left:* CMD of 503113  $z < 0.2$  galaxies taken from the SDSS DR12 where the shading scales with the density of points. The two 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 left. New [O II] luminosity (and subsequently fluxes) are assigned to catalog galaxies from slice sampling the probability histogram.

warning, and place each galaxy on a color-magnitude diagram (CMD) of  $M_r$  and  $g-r$ , see Figure 2.

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 2 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 two colored boxes. The histograms shown in the Figure’s left panel show the probability density histograms of the Log [O II] luminosity for the SDSS galaxies in the 2D bin. We sample the distribution and assign each catalog galaxy an [O II] luminosity which is then converted into a flux.

### 2.3 Mock Observations

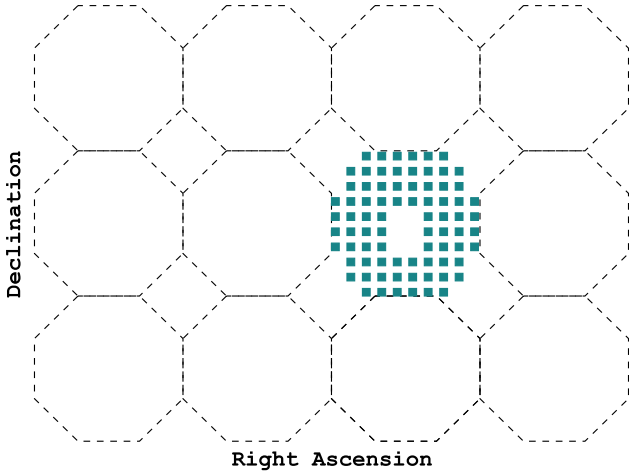
Not sure this does a good enough job talking about the two different observations. Tentatively slated to start in the spring of 2016, HETDEX will perform blind spectroscopy ( $R \sim 750$  in  $3500 - 5500 \text{ \AA}$ ) over two fields along the celestial equator. The  $300 \text{ degree}^2$ , spring field and  $120 \text{ degree}^2$ , 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.

The spectral coverage allows for the detection of [O II] ( $\lambda\lambda 3727 - 3729 \text{ \AA}$  doublet) emitters to  $z \sim 0.5$  and Ca H ( $\lambda 3968.5 \text{ \AA}$ ) and K ( $\lambda 3933.7 \text{ \AA}$ ) absorption features to  $z \sim 0.4$ . HETDEX is expected to detect sources with continuum brighter than 22 mag in  $g$ , and emission line strengths above  $3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ . So we “observe”  $z < 0.4$  galaxies which meet either the emission line or the magnitude limit. Galaxies  $0.4 < z < 0.5$  are only observed if their emission line strength is sufficient. **These limits have been changed. FIX!**

In this work we consider two separate observation cases. The first are targeted observations where we select each galaxy cluster and “observe” each galaxy within  $8'$  of the center. The second is a survey case where observations which are blind to the positions of the clusters are conducted. In both cases, our “observations” consist of placing masks down onto the Buzzard “truth” catalogs and selecting all,  $z < 0.5$  also meeting sensitivity limits, galaxies which lie underneath. Each mask is created to accurately reproduce the HETDEX IFU pattern, see Figure 3. 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.

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



**Figure 3.** 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.

the Buzzard catalog we assume an observation for all galaxies laying within a colored, IFU square in Figure 3. **This should be updated with the fiber collisions.** 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  $375.67 \text{ degree}^2$  field of the Buzzard catalog we require 5370 pointings where  $0.015 \text{ degree}^2$  of each pointing is covered by an IFU. The total area of the sky covered by an IFU is  $80.80 \text{ degree}^2$  which gives a filling factor of  $1/4.65$  slight decreased from the expected filling factor of  $1/4.5$ .

### 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 (LOSVD) 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 (LOSVD) to each galaxy, where

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

and  $c$  is the speed of light in  $\text{km s}^{-1}$ ,  $z$  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

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

where  $\sigma_p$ ,  $\langle u \rangle$ , and  $\sigma_i$  is the LOSVD, the average radial velocity and the error on the individual LOSVDs respectively. Using a Monte Carlo Markov Chain sampler (EMCEE<sup>1</sup>, Foreman-Mackey et al. 2013), we draw twenty thousand samples from the posterior probability distribution. Simple priors,  $\langle u \rangle$  lies between the maximum and minimum LOSVD and  $0 < \sigma_p$ , are used. When the full distribution of LOSVDs are not used, the final LOSVD is quoted as the median value of the 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  $\sigma_p$ . We prefer the maximum likelihood method for its straight forward treatment of the errors in the LOSVD measurements. **Maybe run a quick check to see how they compare.**

#### 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  $\sigma_{15}$  in Evrard et al. 2008 and other works),  $\alpha = 1/3$ ,  $h(z) = H(z)/100$ , and  $\sigma_{1D}$  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

<sup>1</sup> <http://dan.iel.fm/emcee/current/>



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

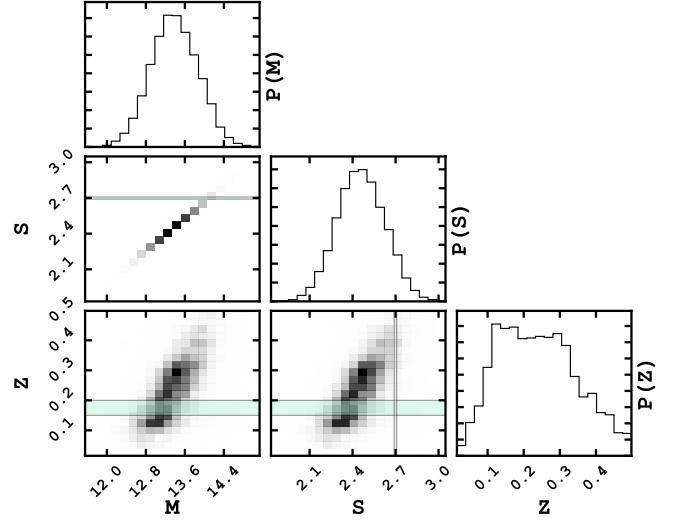
We begin with the training sample. After selecting the desired features  $\vec{x} = \{\sigma, z, \dots\}$  we make the joint probability between the true cluster mass ( $M$ ) and  $\vec{x}$ . Because  $\vec{x}$  can be multidimensional, we rely on the corner plot to visualize the relationship between all of the training features. Figure 4 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  $P(M|\vec{x} = \{x_1, x_2, \dots\})$  is determined by taking a slice through the joint probability distributions in bins centered on the desired value. The slices show by the colored bars in Figure 4 are centered on  $\sigma = 500 \text{ km s}^{-1}$  and  $z = 0.17$ . The distribution of mass contained in the three dimensional bin given by the intersection of these slices is  $P(M|\vec{x} = \{\sigma = 500 \text{ km s}^{-1}, z = 0.17\})$ .

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  $\sigma$  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



**Figure 4.** Corner plot of the *training* data with features  $\sigma$  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 show the slices needed to create a conditional probability distribution of the mass,  $P(M|\vec{x})$ . See text for a complete description.

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 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 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).

The 68% prediction interval is determined by calculating the 16th and 84th percentile of all the predictions made by the ensemble of estimators.

## 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 5 shows the recovery fraction of member galaxies, the number of observed versus the number of actual galaxies, as a function of both redshift and cluster mass for the targeted and survey observing strategies. 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 for nearly all redshifts and cluster masses. This performance gain is due solely to the gaps between the fiber bundles of the IFU.

why does the recovery fraction decrease with redshift? Are the clusters smaller? Do they not have as many galaxies per cluster at higher redshifts? Is it a magnitude effect? I also want to add an error bar estimation for several of the percentages to give a feel for how the different recovery levels effect the accuracy of the measurements.

### 4.2 Mass estimates

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

It doesn't look like you can use the targeted training sample to train the survey data. If you do, it either does worse or about the same if you didn't use any training at all. After looking at it, you do have to use the survey training data to get good results about 0.13 dex MAE

Lorem ipsum dolor sit amet, consectetur adipisicing 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.

## 5 HETDEX AS A GALAXY CLUSTER SURVEY

### 5.1 Expected Results

Lorem ipsum dolor sit amet, consectetur adipisicing 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.

### 5.2 Potential Improvements

Lorem ipsum dolor sit amet, consectetur adipisicing 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.

### 5.3 Extendability to Other Surveys

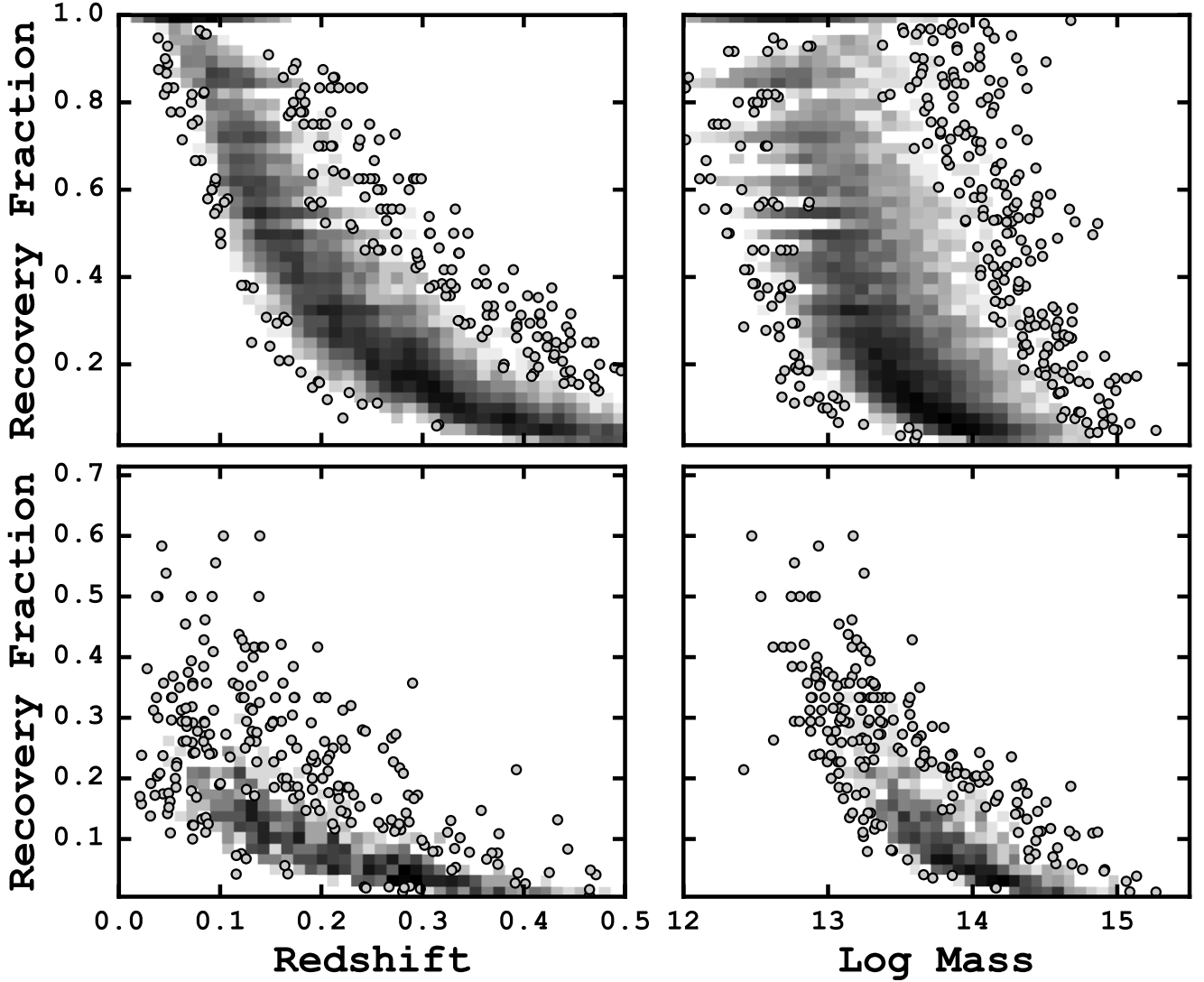
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## 6 SUMMARY

Lorem ipsum dolor sit amet, consectetur adipisicing 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.

## 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 APLPY, an open-source plotting package for Python hosted at <http://aplp.github.com>; the IPYTHON package (Perez & Granger 2007); MATPLOTLIB, a Python library for publication quality graphics (Hunter 2007). 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

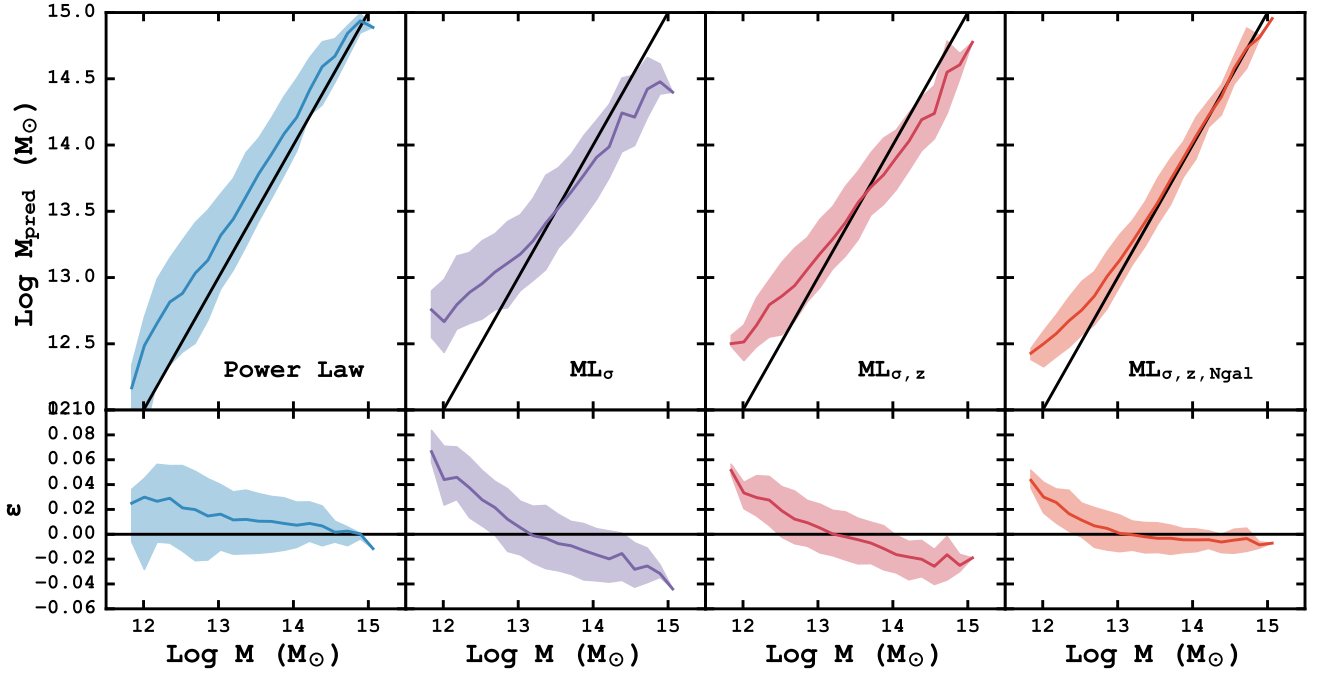


**Figure 5.** Recovery fractions ( $N_{obs}/N_{True}$ ) of cluster member galaxies as a function of redshift and mass for the targeted (upper row) and survey (lower row) observing strategies. The significant decline in galaxies observed with the survey strategy is due to gaps in the VIRUS IFU.

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.

## REFERENCES

- Abell G. O., 1958, *ApJS*, 3, 211  
 Acquaviva V., Gawiser E., Leung A. S., Martin M. R., 2014, *Proc. Int. Astron. Union*, 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, *Publ. Astron. Soc. Pacific*, 123, 568  
 Chabrier G., 2003, *Publ. Astron. Soc. Pacific*, 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, *Publ. Astron. Soc. Pacific*, 125, 306  
 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](https://doi.org/10.1117/12.925434), <http://adsabs.harvard.edu/abs/2012SPIE.8446E..0NH>  
 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](https://doi.org/10.1117/12.2056384), <http://adsabs.harvard.edu/abs/2014SPIE.9147E..75K>  
 Kirk B., et al., 2015, *MNRAS*, 449, 4010



**Figure 6.** Mass predictions for the power law scaling relation (Equation 3) and the ML technique with different input features. The bottom row of panels shows the fractional error (Equation 7) as a function of mass. The solid black line shows the 1:1 relation. The colored lines and shaded region show the median value and 68% bounds respectively.

LSST Dark Energy Science Collaboration 2012, arXiv Prepr. arXiv:1211.0310, 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

Milvang-Jensen B., et al., 2008, *Astron. Astrophys.*, 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, Markov Chain Monte Carlo Methods Based on ‘Slicing’ the Density Function. Department of Statistics, University of Toronto, Toronto, doi:10.1.1.48.886

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, *Astron. Astrophys.*, 571, 19

Ramsey L. W., et al., 1998, in *Proc. SPIE.* pp 34–42, doi:10.1117/12.319287, [http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=945060](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., 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

Sifón C., Hoekstra H., Cacciato M., Viola M., Köhlinger F., van der Burg R. F. J., Sand D. J., Graham M. L., 2015b, *Astron. Astrophys.*, 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, *Astron. Astrophys.*, 561, A79

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