High Precision Galaxy Cluster Mass Estimation: Building a Model for Stacked Caustics in Simulations and its Application to Galaxy Cluster Survey Data

By

Nicholas S. Kern

A Senior Thesis
Submitted in Partial Fulfillment of the
Requirements of Honors BS in the
Department of Astronomy & Astrophysics
University of Michigan

Advisor: Dr. Christopher Miller

April 27$^{\text{th}}$, 2015

Abstract

The true nature behind dark matter and dark energy in our universe are arguably two of the most challenging astrophysical questions of the 21st century. They are so closely connected to the history of our universe that the study of these questions is analogous to the study of the universe's cosmological evolution. The natural laws that govern their behavior will likely only be solved via a multi-faceted approach, where multiple scientific techniques converge upon one coherent picture. One important technique in our scientific toolbox is galaxy cluster cosmology: using the number, size and spatial distribution of galaxy clusters in our universe to trace the growth of structure over cosmic time. A crucial component of galaxy cluster cosmology requires knowing the masses of clusters in the near and far universe. Because these masses are composed mostly of dark matter, determining their total gravitational mass is not easy. One important class of methods to do this is dynamical methods, which use the kinematic positions and velocities of individual galaxies orbiting the cluster's central mass to trace out the cluster's gravitational potential and mass.

In this thesis, we take a close look at one specific dynamical method called the Caustic Technique. We show that the Caustic Technique recovers precise and accurate galaxy cluster masses in the limit where we have spectroscopic redshifts for a large number of galaxies per galaxy cluster ($N_{gal} > 75$). However, even with the millions of galaxy spectra already taken by current galaxy cluster surveys, for the majority of known galaxy clusters in the universe we typically have less than 25 galaxy spectra per cluster, meaning we cannot achieve accurate or precise masses for these clusters with the Caustic Technique–or any dynamical method for that matter. In order to circumvent this problem, we developed a stacking algorithm that combines the galaxy data from multiple galaxy clusters into one ensemble galaxy cluster, which increases the galaxy sampling to where we expect to be unaffected by low-number statistics. We use large N-Body simulations to systematically test the performance of our algorithm and determine its uncertainties. We show that we can recover accurate and precise average galaxy cluster masses from clusters that have as little as 10 galaxy spectra per cluster. We then apply our technique to a real spectroscopic sample and characterize its Mass–Richness relationship. We self-calibrate the relationship using our stacking algorithm and produce a corrected relationship that is in agreement with our expectation from simulations and is unbiased with respect to galaxy cluster sampling uncertainties. Proof that our stacking algorithm is not only feasible for use in the real universe but advantageous in certain scenarios opens the door for scientists to use dynamical mass estimators to do cosmology with incredibly large datasets of galaxy clusters that were before effectively unusable for such purposes.

Acknowledgements

First, I am extremely thankful for the academic guidance and support from Professor Chris Miller and PhD candidate Dan Gifford at the University of Michigan. During my time as an undergraduate at Michigan they took me under their wing and played an instrumental role in shaping my scientific mind and fostering my desire to pursue astrophysics research as a career. I would also like to express my gratitude toward my two academic advisors at Michigan, Ted Bergin and Nuria Calvet, who have created a fertile environment for undergraduates to engage in research and academically thrive. I also want to thank all of the people associated with the Astronomy and Physics departments–faculty, researchers, staff, graduate students and my fellow undergrads–for being a part of my undergraduate educational experience that has profoundly shaped me as a person. Last, but perhaps most importantly, I want to thank and acknowledge my family and friends, who have always supported me in my life and have provided me with a platform of stability from which I have been able to pursue my academics, which I often forget is an incredible privilege.

Contents

List of Figures

1. Introduction

Galaxies in our universe are not distributed evenly across the sky. When we point our telescopes at any region of the sky, we see that galaxies are highly clustered in some regions and devoid in others. This bunching of galaxies is due to gravitational instability, which pulls matter together to form clusters of galaxies. These large ($\sim 10^6$ pc) systems are called galaxy clusters–sometimes referred to as just clusters or dark matter halos–and are the largest gravitationally bound objects in the universe.

Galaxy clusters are not bound by the gravity from the stars and dust in individual galaxies. It was shown in the 1930s that the Coma cluster, for example, had a mass inferred from it gravity that was a few hundred times the mass inferred from the starlight from all the galaxies in the cluster. A component of mass invisible to optical light was proposed as a possible solution to the problem, called dark matter (Zwicky 1937). Although it was later found that the medium between galaxies is filled with a hot gas at millions of degrees that makes up most of the baryonic matter of galaxy clusters, an additional mass component was still needed to bridge the gap between virial and baryonic mass estimates. Today, we still call this missing mass dark matter, and its mysterious origin is one of the questions at the forefront of modern astrophysics.

It is now known that dark matter is the dominant matter component of the universe, comprising over 85% of its total matter content (Komatsu et al. 2011). This means that the normal baryonic matter–that which makes up our everyday lives–is by far a subdominant component of the universe. This is no exception in galaxy clusters: when we talk about the mass of a galaxy cluster, we are speaking of its dark matter halo. In fact, all galaxy cluster mass estimation techniques are merely ways through which we can measure the presence of a dark matter halo through its gravitational interaction with light or baryonic matter, which we can detect.

The current picture of galaxy cluster formation starts with the Big Bang, where quantum fluctuations in the distribution of matter and radiation in the hot early universe seeded over-dense and under-dense regions. To first order, the over-dense regions grew linearly over time and pulled more and more matter into their gravitational potential wells. Over time these potential wells continued to grow and eventually developed into the deep potential wells that make up the galaxy clusters we see today, 13.8 billion years after the Big Bang. The study of galaxy clusters is, then, the study of the growth of large-scale structure in the universe. The growth of structure in the universe is strongly correlated to the underlying cosmology (e.g. Tinker et al. 2008). Knowing the distribution, size and number of galaxy clusters as a function of time is therefore a unique probe that can constrain cosmological models that govern the past and future evolution of our universe.

In this thesis, I explain the research I have done over the past four years to better understand the dynamics of galaxy clusters and use them as cosmological probes. The single, most important endeavor that my research is concerned with is calculating the total Newtonian mass of a galaxy cluster. Specifically, I take a close look at one particular dynamical method called the Caustic Technique. I show that the Caustic Technique recovers precise and accurate galaxy cluster masses in the limit where we have spectroscopic redshifts for a large number of galaxies per galaxy cluster ($N_{gal} > 75$). However, even with the millions of galaxy spectra already taken by current galaxy cluster surveys, for the majority of known galaxy clusters in the universe we typically have less than 25 galaxy spectra per cluster, meaning we cannot achieve accurate or precise masses for these clusters with the Caustic Technique–or any dynamical method for that matter. In order to circumvent this problem, I co-developed a stacking algorithm that combines the galaxy data from multiple galaxy clusters into one ensemble galaxy cluster, which increases the galaxy sampling to where we expect to be unaffected by low-number statistics. I use large N-Body simulations to systematically test the performance of the algorithm and determine its uncertainties. I show that it recovers accurate and precise average galaxy cluster masses from clusters that have as little as 10 galaxy spectra per cluster. I then apply the technique to a real spectroscopic sample and characterize its Mass–Richness relationship. I self-calibrate the relationship using the stacking algorithm and produce a corrected relationship that is in agreement with the expectation from simulations and is unbiased with respect to galaxy cluster sampling uncertainties. Proof that the stacking algorithm is not only feasible for use in the real universe but advantageous in certain scenarios opens the door for scientists to use dynamical mass estimators to do cosmology with incredibly large datasets of galaxy clusters that were before effectively unusable.

The sectioning of this document begins with a general discussion of how we derive the masses of galaxy clusters and details a systematic test of the Caustic Technique in §2. Much of chapter 2 has been published in Gifford, Miller, & Kern (2013). In §3, I explain the theory behind the stacking algorithm and describe our rigorous and systematic tests in simulations to understand its uncertainties. The results of this chapter are currently being submitted to the Astrophysical Journal (Gifford, Kern, Miller et al. 2015). Finally, in §4, I present the stacking algorithm's application to real galaxy cluster optical survey data and discuss its implications for future surveys. The results of this chapter are also currently being submitted to the Astrophysical Journal (Kern, Gifford, Miller et al. 2015). The techniques I developed in §3 and the algorithms I used to apply them to real data in §4 are also being used in related work concerning modified gravity in clusters (e.g., Stark, Miller, Kern et al. 2015).

2. Deriving the Masses of Galaxy Clusters

2.1. Introduction

Knowing the masses of galaxy clusters is important to cosmology because any theoretical model of structure formation must be able to reproduce the observed mass function: the number of clusters in the universe in a certain mass range. Cluster masses also provide the basis for a variety of mass-to-observable scaling laws, which are relationships that correlate a cluster's dark matter mass to some observable signature of the cluster, such as its luminosity, X-ray temperature, galaxy richness or the velocity dispersion of its orbiting galaxies. In many cases, observing these mass proxies are quicker and easier than actually deriving the mass of a cluster via some measure of its gravitational potential. This means that it becomes feasible to measure these mass proxies over large portions of the sky and trace how these relationships change over the history of the universe. If we can accurately calibrate these scaling laws with a subsample of highly precise cluster masses, we will be able to statistically determine the masses of clusters in large-scale surveys and constrain cosmological models.

2.2. How Do We Define A Mass?

Deriving the masses of clusters, however, is not an easy process. In some sense it is not even clearly defined. When does a dark matter halo effectively terminate, if it does at all? The popular Navarro-Frenk-White dark matter density model shown to match N-Body dark matter density profiles, for example, does not converge when integrated to infinity (Navarro et al. 1996). To what radius should we integrate a mass density profile to derive a mass? One way of choosing a radius is by utilizing the virial theorem, which states that a "relaxed" gravitational system will have a total kinetic energy equal to one half of its gravitational potential energy

$$T_{\text{kinetic}} = -\frac{1}{2}\Phi_{\text{gravity}} \tag{1}$$

In this case, "relaxed" means that the galaxies in a cluster are gravitationally bound and have spent enough time in the cluster's gravitational well that their orbits reflect the influence of the cluster's gravitational potential. In other words, the contribution of their relic infall velocities to their overall instantaneous velocities are negligible compared to the velocities induced by the gravitational potential of the cluster.

This condition is satisfied in the cores of most galaxy clusters, but will be broken at some radius from the cluster center, denoted as r_{vir}. This radius also happens to be approximately equal to r_{200}, the radius at which the cluster's mass density, ρ, is approximately equal to 200 times the critical mass density of the universe ρ_{crit}. This critical density is derived from the Friedman equations and can be written as $3H^2/8\pi G$, where H is the Hubble constant and quantifies the expansion rate of the Universe and G is the Newtonian gravitational constant. ρ_{crit} is a constant 9.2×10^{-27} kg/m^3 (Ryden 2003).

Gravitational potential is related to the mass density through the Poisson equation

$$\nabla^2\Phi(r) = 4\pi G\rho(r) \tag{2}$$

If we integrate a cluster's mass density profile out to a radius of r_{200} then we get the mass enclosed in this sphere, M_{200}.

$$\int_0^{r_{200}} \rho(r)dr = M(< r_{200}) = M_{200}$$

This is a common way to define a cluster mass, as the integral of its density profile out to a radius at which the cluster's density is some number times the critical density of the universe. Similarly, we can define M_{500} and M_{1500}, which are the masses enclosed by r_{500} and r_{1500}, which are at significantly smaller radii than r_{200}. In this thesis when we refer to a cluster mass, we are referring to it's M_{200}, or the mass enclosed within a spherical radius where the cluster's density is equal to $200 \cdot \rho_{\rm crit}$.

2.3. How To Do It: The Caustic Technique

There are a few ways to directly measure the gravitational potential of a cluster and thus estimate its mass. One way is to use a galaxy cluster's dark matter halo as a gravitational lens, which bends the light of background galaxies into our line of sight. The magnitude of the bending is a prediction of general relativity and determines the mass of the cluster (e.g. Rozo et al. 2009). Another method uses the microwave emission from the cluster's intracluster medium (ICM), which is a result of inverse Compton scattering of photons from the cosmic microwave background (CMB) off of highly energetic electrons in the cluster (e.g. Planck Collaboration et al. 2014). Another method uses the temperature of X-ray emission from the ICM to back out the gravitational potential and hydrostatic mass of the cluster (e.g. Kravtsov et al. 2006). Yet another method uses the motions of galaxies around the center, defined by simple Newtonian dynamics, to trace out the cluster's gravitational potential and Newtonian mass (e.g. Evrard et al. 2008; Saro et al. 2013; Gifford & Miller 2013). Each technique is subject to various systematic uncertainties, making some preferable over others in various situations. A clear picture of galaxy cluster formation requires a statistically coherent picture of galaxy cluster masses from each individual technique. To date, this has yet to be done.

Our focus is on the latter of the above stated methods, which represents a class of techniques called dynamical methods. A dynamical method simply uses the positions and velocities of galaxies around a cluster to measure the cluster's mass. Within the scope of dynamical techniques, we focus on the **Caustic technique**, originally conceived by Diaferio & Geller (1997) and expanded by Diaferio (1999).

The Caustic technique takes the positions and velocities of a galaxy cluster's member galaxies and projects them into a phase space of cluster-centric radius and line-of-sight velocity (Figure 1). These phase spaces are also referred to as radius-velocity or r-v phase spaces. If we were to create the phase space of a galaxy cluster in three dimensions–taking each galaxy's total radius from the cluster center and it's total velocity vector–we would find that the galaxies in a cluster fall below

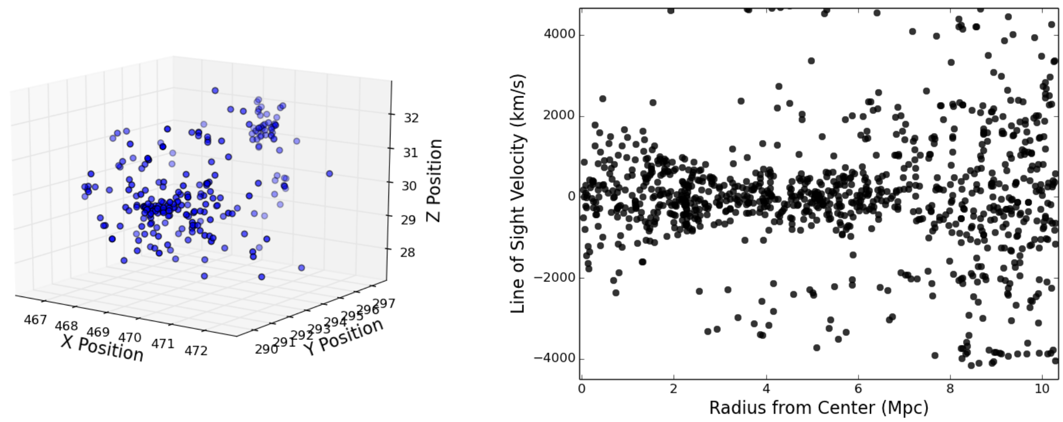


Fig. 1.— **Left:** Simulated galaxy cluster in three dimensional space. **Right:** A phase space of the same simulated galaxy cluster, where individual galaxies' cluster-centric radius and line of sight velocity are projected along a line of sight towards the cluster from Earth. The characteristic trumpet shape seen at small radii in the phase space to the right is the galaxy cluster signal that the Caustic technique searches for.

the escape velocity profile of the cluster. This is a natural consequence of Newtonian dynamics: if a galaxy has a velocity larger than the escape velocity at its radius then it will escape from the cluster, and we should find virtually no galaxies above this threshold (Figure 2). Therefore, the "edge" of the phase space, called the Caustic surface (blue line in Figure 2), traces the escape velocity profile of the cluster, which is directly related to the cluster's gravitational potential via Newtonian dynamics through

$$v_{esc}^2 = -2\Phi(r) \tag{3}$$

This means that if we can accurately trace the escape velocity profile of the cluster then we can directly back out its potential, which according to the Poisson equation (Equation 2) is directly related to its density profile and therefore mass. Combining the virial equation with Equation 3 we get that the escape velocity is related to the average velocity at the virial radius

$$\langle v_{\rm esc}^2 \rangle - 4\langle v^2 \rangle = 0 \tag{4}$$

This process, however, is complicated due to the fact that in the real universe we are viewing a galaxy cluster in projection. We are therefore actually measuring the line-of-sight escape velocity profile $v_{\rm esc,los}(r)$ and the line-of-sight galaxy velocities $v_{\rm los}$. The anisotropy of the galaxies' velocity is encompassed by an anisotropy parameter β, which consists of a galaxy's tangential velocity v_θ and radial velocity v_r

$$\beta(r) = 1 - \frac{\langle v_\theta^2 \rangle(r)}{\langle v_r^2 \rangle(r)} \tag{5}$$

Diaferio (1999) shows that in projected space, the line of sight escape velocity is related to the true escape velocity using β(r):

$$\langle v_{\rm los,esc}^2 \rangle(r) \approx \frac{1-\beta(r)}{3-2\beta(r)} \langle v_{\rm esc}^2 \rangle(r) = \frac{1}{{\rm g}(\beta(r))} \langle v_{\rm esc}^2 \rangle(r) \tag{6}$$

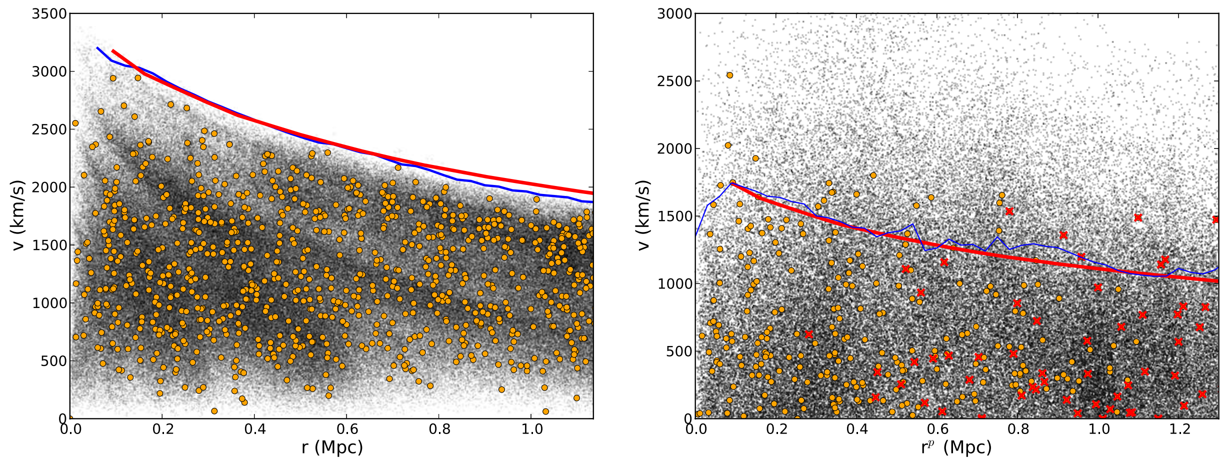


Fig. 2.— **Left:** Simulated three dimensional phase space of a galaxy cluster. **Right:** Simulated projected phase space of the same galaxy cluster along its line of sight towards us. In both figures, the black points represent dark matter particles, while the orange dots represent simulated galaxies. The red line represents the cluster's gravitational potential, while the blue line is the estimate caustic surface. Red crosses in the right figure are interloper galaxies projected into the phase space. Figure taken from (Gifford et al. 2013).

We therefore use g(β) to transform our projected escape velocity to a three dimensional escape velocity.

We now have an estimate of the escape velocity and hence gravitational potential via Equation 3. One course of action to get the cluster's mass is to use the Poisson equation (Equation 2) to relate potential profile to a mass density profile. This, however, involves taking a derivative of a non-smooth function. Diaferio (1999) introduces an alternative method using a partial mass differential equation $dm = 4\pi\rho(r)r^2dr$. If we invoke Equation 1, this becomes:

$$dm = -2\pi v_{\rm esc}^2(r)\frac{\rho(r)r^2}{\Phi(r)}dr \tag{7}$$

We can multiply this by the gravitational constant and integrate to yield the mass inclosed by some radius R:

$$\begin{aligned} G\cdot M(<R) &= \int_0^R -2\pi\cdot G\cdot g(\beta(r))\cdot\langle v_{\rm esc,los}^2\rangle(r)\cdot\frac{\rho(r)r^2}{\Phi(r)}\cdot dr \\ &= \int_0^R \mathcal{F}_\beta(r)\cdot\langle v_{\rm esc,los}^2\rangle(r)\cdot dr \end{aligned} \tag{8}$$

where the term $\mathcal{F}_\beta(r)$ encompasses the $g(\beta(r))$, $\rho(r)$ and $\Phi(r)$ terms. Diaferio (1999) claim that $\mathcal{F}_\beta(r)$ is roughly constant as a function of radius from 1 to 3 times r_{200}, which they calibrate against N-Body simulations. They claim an average $\mathcal{F}_\beta$ of 0.5, while Serra et al. (2011) claim an average of 0.7. Differences in these values are related to the kind of mass tracers used in one's mock catalogues, such as galaxies, dark matter particles or subhalos. While Diaferio (1999) used subhalos as velocity tracers–which are known to be biased tracers of the velocity dispersion–Serra et al. (2011) used dark matter particles. For our analysis, we adopt an $\mathcal{F}_\beta$ of 0.65, which is calculated using galaxy tracers explained in subsection 2.4.

Going from a phase space to an escape velocity profile is not so simple, however. Once a phase space is constructed, the Caustic technique runs a two dimensional kernel density algorithm that creates a density space overtop of the phase space. The technique then picks a set of iso-density contours from the density space that outline the edge of the galaxies in the phase space, which generally forms a characteristic trumpet shape (Figure 1 Right). The next question is, which iso-density contour best matches the escape velocity profile (see Figure 2)? We do this two different ways. One way we do this is by taking the iso-density contour that best fits the velocity dispersion of the cluster, thereby making the technique dependent on the cluster's velocity dispersion. This is what we typically associate with a Caustic mass. Another way, which is currently being developed and tested by our team, is to bin the data in radial bins and take the top X percent ($\sim$10%) of galaxies along the velocity axis in each bin, thereby tracing out a Caustic surface as a function of radius. We call a mass derived with this variant of the Caustic technique an Edge mass.

When working on individual galaxy clusters–i.e. phase spaces of individual galaxy clusters–the typical Caustic masses are preferred. When we start working on stacked phase spaces, as is described in section 3, we see that we want to use an Edge mass.

2.4. Systematic Analysis of the Caustic Technique

In 2013, we presented a systematic analysis of the Caustic technique against simulated dark matter halos (Gifford, Miller, & Kern 2013). Simulations provide an ideal proving ground for testing the performance of the technique because we know what the *real mass* of a cluster is in a simulation: we just physically add up the masses of all the dark matter particles within r_{200}. This means we can test the technique, change it, and test it again to optimize it's performance. To do this, we queried a set of 100 dark matter halos from the Millennium Simulation, a large simulation that models the gravitational clustering of dark matter particles over cosmic time(Springel et al. 2005). The Caustic technique, however, relies on the motions of galaxies, not dark matter particles. We therefore used a suite of semi-analytic catalogues–algorithms that paste galaxies onto dark matter subhalos and model galaxy formation processes–to test the performance of the Caustic technique against the simulated halos and their galaxies (Springel et al. 2001; Bower et al. 2006; De Lucia & Blaizot 2007; Bertone et al. 2007; Guo et al. 2011). We discuss these simulations further in subsection 3.2.

One of the results of the study was that a shiftgapper method of interloper treatment works well for dealing with interloper galaxies in a phase space, better than a sigma clipping routine. A shiftgapper method bins the phase space data in radial bins and measure the velocity dispersion of galaxies in each bin. In each bin it looks for velocity gaps where there are no galaxies, and if the velocity gap is bigger than some threshold it cuts all galaxies above the gap. We will continue to use a shiftgapper technique to deal with galaxy interlopers for the rest of this thesis.

The other main result we are interested in is how the Caustic technique performs in a simulation that is made to mock a real observation. In such a scenario, we choose a galaxy cluster in three dimensional coordinates and put ourselves some distance away from the known cluster center. We then project every galaxy into a two dimensional space, identical to how we would see it in the sky, and make realistic cuts in galaxy magnitude and completeness. Using four different semi-analytic datasets and the dark matter subhalos, Gifford et al. (2013) finds that the Caustic technique is fairly robust to semi-analytic modeling. Figure 3 shows that the Guo et al. (2011), Bertone et al. (2007) and De Lucia & Blaizot (2007) catalogues give essentially identical virial mass, Caustic mass and velocity dispersion biases as a function of galaxy sampling number, N_{gal}. The subhalos and the Bower et al. (2006) semianalytics are distinctly different than the other catalogues in that they treat orphan subhalos/galaxies differently, which is likely the cause of their mass bias. What is also important, is that when a phase space is sampled with roughly 50 galaxies or more the Caustic returns unbiased masses and has a scatter of about 35% (Figure 3 & Figure 4). The phase space sampling, i.e. the number of galaxies in a phase space, is key to how the performance of the Caustic and is denoted by $N_{gal} \equiv$ "number of galaxies in a phase space."

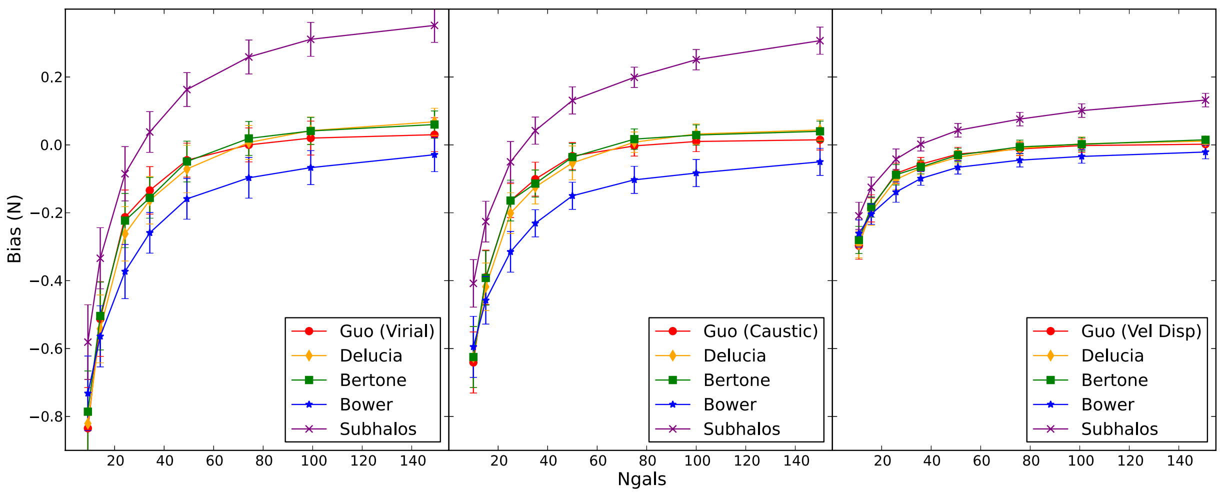


Fig. 3.— The Caustic technique is fairly robust to how galaxies are modeled in simulations. In the middle panel, three out of four semi-analytic catalogues give essentially identical results: when the phase space is sampled with > 50 galaxies ($N_{gal} > 50$) the Caustic technique returns unbiased masses with respect to the Simulation's quoted halo mass. Figure from Gifford et al. (2013).

2.5. Conclusion: Motivation for Stacking

While it is reassuring that the Caustic technique can theoretically produce unbiased ($0 \pm 5\%$) and low scatter (∼30%) halo mass estimates, this is contingent upon having at least 50 member galaxies in our phase space. Observationally, this means getting at least 50 spectroscopic redshifts per cluster, which is feasible and available for mid and high mass clusters at low redshift. However, this gets increasingly more difficult for low-mass clusters and for all clusters at high redshift. In fact, all dynamical mass estimators depend heavily on having a statistically significant sample of N_{gal} measured galaxy redshifts. Individual clusters that have less than ∼30 galaxy spectra are therefore rendered effectively useless for dynamical mass estimators.

This is somewhat disconcerting, especially because for any large scale galaxy survey, the clusters on the edge of its redshift limits will suffer from poor galaxy sampling due to magnitude limitations. This, however, is exactly the region that we would like to probe for cluster masses to

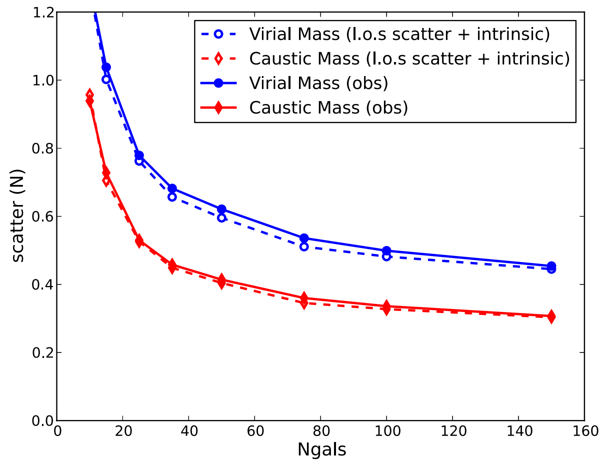


Fig. 4.— The Caustic technique has a fundamental floor on how precise it can be (red). It levels out at about 30% – 35% scatter for very high phase space sampling ($N_{gal} > 100$). This is largely due to projection effects, and clusters' spherical asymmetry. Figure from Gifford et al. (2013).

do exciting cosmology: the bleeding edge of cluster surveys at higher and higher redshifts. This will be particularly important over the next decade when (example of surveys) deliver unprecedented quantities of optical data that can be used to constrain the cosmological growth of dark matter. Some of the most interesting clusters in these surveys will fall near the magnitude limits of the telescopes and will thus be too sparsely sampled to be used with conventional dynamical mass estimators.

This provides a strong incentive to develop a method that will allow dynamical techniques to at least extract some useful information from these poorly sampled clusters. One way to circumvent this problem is to combine the sparsely sampled data sets together in order to increase the total signal of a resultant data set. This only makes physical sense if the data sets are drawn from the same or a similar underlying distribution. In our case, this would be the underlying dark matter mass profile of each galaxy cluster. Assuming we take galaxy clusters that are similar in size and mass, if we ***stack*** the phase spaces of their member galaxies together, we may be able to make a more precise measurement of their ***total averaged mass***.

This kind of technique is actually quite common in astronomy. It is has, for example, already been implemented in the field of galaxy cluster cosmology through stacked weak lensing shear methods. This technique uses the bending of light around galaxy clusters as predicted by general relativity to trace the mass profile of the central cluster. In many cases the bending of the light or "shear" is statistically too weak to derive a precise measurement of the central cluster's mass. When these shear measurements are stacked overtop many other weakly lensed shear data, a coherent picture emerges. In a similar way, we thought that by stacking individual galaxy clusters into an ensemble cluster, we might be able to measure a Caustic surface and produce an average Caustic mass for the ensemble. There have been studies that have stacked galaxy cluster kinematic data to study the non-mass related dynamics of ensemble clusters, such as their velocity dispersions (Becker et al. 2007). However, it has yet to be investigated whether one can derive a reliable dynamical mass estimate from stacked ensembles, and if those can then be used to constrain cosmology. We therefore set out to build a model for such a technique, with the overarching goal of eventually applying it to real data if we could reasonably understand the systematic errors associated with stacked galaxy cluster phase spaces.

3. Stacked Caustics: Towards Precise Dynamical Masses

3.1. Introduction

It is not immediately clear that stacking is possible with dynamical mass estimators: part of our work was to figure out if and when applying a stacking method is appropriate. In order to create a model for stacked galaxy clusters, we developed and tested our algorithms against a wealth of N-body and semi-analytic simulation data from the Millennium Simulation.

When we "stack" galaxy clusters, what we mean is that we are overlaying the cluster-centric radius and line of sight velocity phase spaces of multiple clusters on top of each other to create a master phase space. This really only makes sense if the galaxies from each cluster are drawn from a similar distribution. This corresponds to taking galaxy clusters that are similar in mass, or some observable correlated to mass such as richness or luminosity, and combining them to create an *ensemble cluster*. This has the direct effect of increasing the sample size of galaxies in our phase space, which means that we can measure the Caustic surface to higher precision.

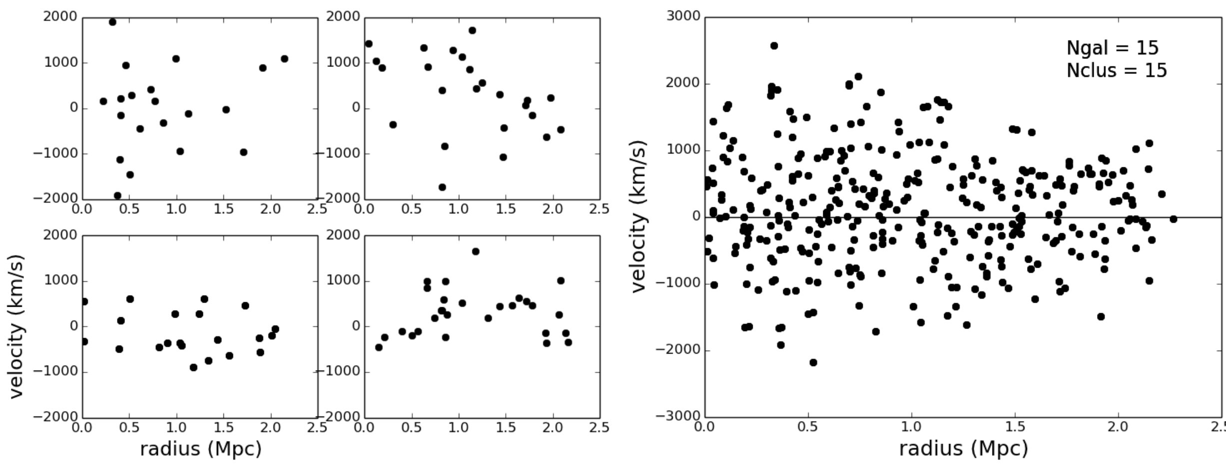


Fig. 5.— **Left:** Four sparsely sampled individual galaxy clusters from MS with N_{gal} = 15 for each cluster. In some cases the galaxies extend to ± 2000 km/s and in others only ± 1000 km/s, which is in part due to projection effects. **Right:** Stacked ensemble cluster including clusters at left with N_{clus} = 15 and an N_{gal} = 15 for each stacked cluster. This equates to an ensemble sampling of $\sim N_{clus} \cdot N_{gal} = 225$ galaxies within the ensemble cluster's r_{200}.

Stacking also has the added benefit of averaging out projection effects, which adversely affects the Caustic technique's performance because it assumes spherical symmetry. Galaxy clusters are not completely spherical and can resemble ellipsoids to some degree. Because clusters are randomly oriented in space, we will sometimes view an ellipsoidal cluster on its side, as if we were looking at the laces of a football, and sometimes we will view it down the barrel, as if we were looking straight at the tip of a football. These differences cause discrepancy in the determination of the cluster's velocity dispersion and will decrease or increase the Caustic surface respectively, however, when we combine multiple randomly oriented systems these differences are effectively averaged out (Figure 5). This large source of uncertainty that affects the Caustic technique's precision on individual clusters is actually alleviated in stacked ensemble clusters, such that stacked Caustic masses theoretically level out at much higher precision (Figure 8). Here, we present the first viable model for dynamical mass estimation of stacked galaxy cluster phase spaces. We show that this technique can theoretically decrease the scatter in Caustic mass estimates from $\sim 35\%$ to as low as $\sim$10%.

3.2. Developing the Stacking Technique: Simulation Data, Computational Challenges and Methodology

3.2.1. Data

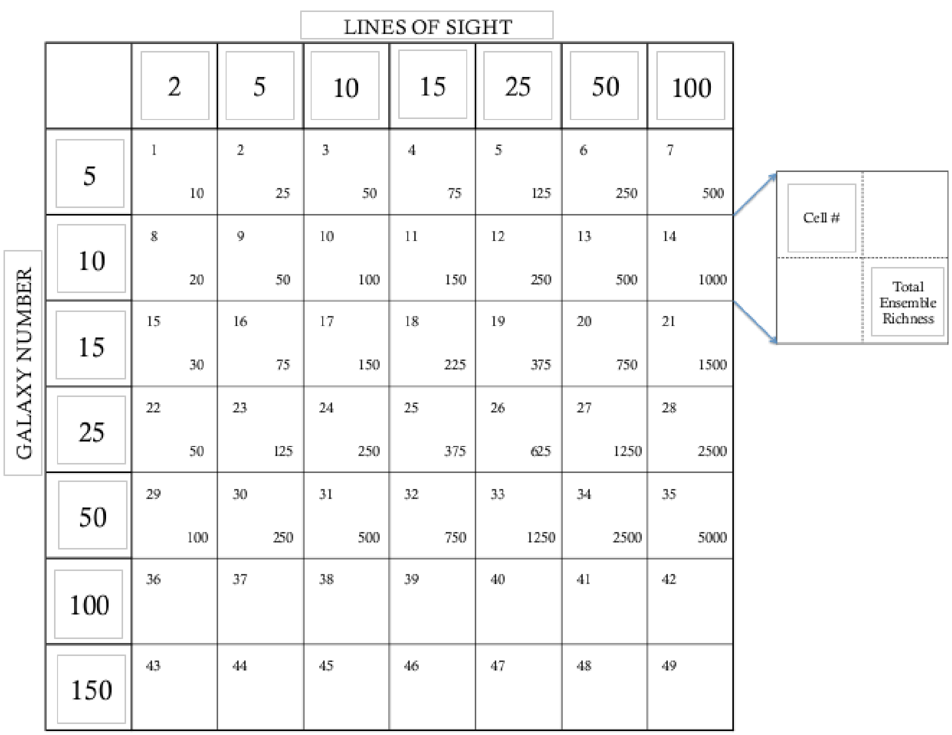
The N-Body simulation data that we use comes from the Millennium Simulation, hereafter MS, developed and run by the VIRGO Consortium, a collaboration of British, German, Canadian and US astrophysicists (Springel et al. 2005). They utilized the open sourced GADGET2 code (Springel et al. 2001) to carry out the simulation and calculate the gravitational forces between dark matter particles over time. The cosmological parameters for their simulation follow the standard ΛCDM model with cosmological parameters $\Omega_M = \Omega_{DM} + \Omega_b = 0.25, \Omega_b = 0.045, h = 0.73, \Omega_\Lambda = 0.75, n = 1, \sigma_8 = 0.9$, and $H_0 = 100 \cdot h$ km s^{-1} Mpc^{-1}. At the time it was run in 2005, the MS was the largest ever N-Body simulation of cold dark matter by a factor of 10. The MS itself monitors *only the dark matter particles* and uses a Friends-Of-Friends (FOF) algorithm to determine when a clustering of dark matter particles constitutes a "subhalo," and when a clustering of subhalos constitutes a "halo." To develop the stacking technique, we extracted data for over 2,000 dark matter halos from their database–a factor of 20 greater than our previous analysis in subsection 2.4.

On top of the motions of the dark matter subhalos, galaxy formation, evolution and destruction can be characterized and mapped over time using semi-analytic models that impose basic physical rules on the star formation, radiative transfer, AGN feedback and supermassive black hole growth for each galaxy. These are called semi-analytic catalogues, and in our study we utilized the Guo et al. (2011) semi-analytic catalogue to extract the motions of galaxies around each dark matter halo in the MS. The Guo semi-analytic model is tuned so that the results at the end of the simulation match the observed galaxy population in the local universe. We can therefore test our code in simulations and make realistic predictions as to how our technique will perform in the real universe.

The Guo catalogue, however, is a data cube. This means that the position and velocity vector of each galaxy is defined in a three dimensional cartesian coordinate frame in X, Y and Z with respect to the simulation box. In the real universe we work in sky coordinates of Right Ascension (RA) and Declination (Dec) for each galaxy's position and redshift (z) for each galaxy's line of sight velocity. To model a mock observation in the MS, Henriques et al. (2012) created a "light cone" out of the Guo semi-analytic catalogue. This means that they took the Guo simulation box at different times during its evolution–at a z of 0, 0.5, 1, 1.5, 2 and so on–and stitched them end to end from an observer to make a model universe as we would see it from Earth. We call this dataset the Henriques lightcone. We use the Henriques lightcone as an intermediary step between developing our technique with the Guo dataset and applying our technique to data in the real universe to ensure we can account for all of the systematic uncertainties of the technique.

3.2.2. Computational Challenges

The way in which we develop and test the stacking technique is highly systematic. When we stack galaxy clusters together there are essentially *two parameters* that affect the dynamics of the resulting ensemble cluster, N_{gal} and N_{clus}. N_{gal} is the number of galaxies we take from each individual cluster and put into the stack, while N_{clus} is the number of individual galaxy clusters that we do this to in order to create one ensemble cluster. To create the ensemble cluster in Figure 5, for example, we took fifteen galaxy clusters and from each, took fifteen galaxies and threw them into the ensemble cluster, meaning that the ensemble cluster has a total sampling size of $N_{gal}{\cdot}N_{clus} = 15 \cdot 15 = 225$. For any one pairing of N_{gal} and N_{clus}, we stack across our full sample of 2000 galaxy clusters. Therefore, the number of resultant ensemble clusters is uniquely defined by N_{clus}, in that $N_{ens} = 2000\ /\ N_{clus}$. We then repeat this for multiple pairs of N_{gal} and N_{clus} to explore how different ways of building an ensemble cluster affects our ability to recover its average mass. This creates a two dimensional parameter space that manifests as a grid, which we call an *ensemble grid*, shown in Figure 6.



	2	5	10	15	25	50	100
5	1 / 10	2 / 25	3 / 50	4 / 75	5 / 125	6 / 250	7 / 500
10	8 / 20	9 / 50	10 / 100	11 / 150	12 / 250	13 / 500	14 / 1000
15	15 / 30	16 / 75	17 / 150	18 / 225	19 / 375	20 / 750	21 / 1500
25	22 / 50	23 / 125	24 / 250	25 / 375	26 / 625	27 / 1250	28 / 2500
50	29 / 100	30 / 250	31 / 500	32 / 750	33 / 1250	34 / 2500	35 / 5000
100	36	37	38	39	40	41	42
150	43	44	45	46	47	48	49

Fig. 6.— Our ensemble grid, describing a two dimensional parameter space where we explore different ways to construct an ensemble cluster by varying N_{gal} (vertical axis) and N_{clus} (horizontal axis). In each cell, the upper left number uniquely identifies the cell and the lower right number is $N_{gal}{\cdot}N_{clus}$.

This creates a serious computational challenge; for each of the 49 cells in our ensemble grid we are stacking our full sample of 2000 galaxy clusters and running the Caustic technique over the resultant ensembles. By itself, the Caustic technique takes about four seconds to calculate the mass of one galaxy cluster phase space. This means that each full realization of an ensemble grid takes a little over 14 hours. However, we need to iterate the ensemble grid many times to average out stochastic processes and constrain the errorbars on our mass estimates. All of this boils down to about one week worth of computation in order to fully test our technique. In order to improve our technique, however, we need to repeat this whole process every time we significantly change it. This high volume of computation time–let alone the large amounts of data we are constantly loading in and writing out–makes developing the code infeasible with a single computer. We therefore moved to High Performance Computing (HPC) and parallel processing to make this possible. In fact, we specifically designed the code in order to be optimized for parallelization and usage on HPC systems.

At Michigan, we have 40 dedicated computers and large amounts of data storage that we can remotely access. In order to get even more computing power, we applied for and won a science allocation of 36 computers on the Open Cloud Consortium's Open Science Data Cloud (OSDC) computing cluster. We remotely configured the 36-core supercomputer and uploaded our data and software in order to leverage a combined 76 computers between our research group. This cut down our one-week computation time by about a factor of 70, which was key in allowing us to proceed with the systematic development of the stacking technique.

In order to promote scientific transparency and allow other researchers to duplicate our results and use our code, we have open sourced our custom built stacking code to the public, which we call "caustic_stack."[1]

3.2.3. Methodology

The notion of stacking introduces inherent uncertainties that we need to be able to understand and control in order for stacking to become a usable technique on real data. The total uncertainty can be thought of as the product of multiple nested probabilities that each describe different uncertainties associated with the technique. Our ultimate goal is to describe the uncertainty of the Caustic mass with respect to the "True" mass of the cluster: $\mathrm{P}(\mathrm{M}_{\mathrm{caustic}}|\mathrm{M}_{\mathrm{true}})$. However, there is some ambiguity in how we choose what the "True" mass of the cluster is, in particular for the case of an ensemble cluster, which is composed of multiple individual clusters where we could reasonably take either a mean, median or weighted mean of the individual masses to derive the ensemble's "True" mass. We therefore represent this ambiguity with an uncertainty variable $\mathrm{P}(\mathrm{M}_{\mathrm{ensemble}}|\mathrm{M}_{\mathrm{true}})$. In the case of the real universe, we don't even know what $\mathrm{M}_{\mathrm{true}}$ really is, so there is an additional uncertainty in the mass proxy that we use to determine clusters that are similar in size, defined as $\mathrm{P}(\mathrm{M}_{\mathrm{observable}}|\mathrm{M}_{\mathrm{true}})$. This gives us a total uncertainty equation:

$$P(M_{\mathrm{caustic}}|M_{\mathrm{true}}) = P(M_{\mathrm{caustic}}|M_{\mathrm{ensemble}}) \cdot P(M_{\mathrm{ensemble}}|M_{\mathrm{observable}}) \cdot P(M_{\mathrm{observable}}|M_{\mathrm{true}}) \qquad (9)$$

[1]The open sourced Python code can be found here: https://github.com/nkern/caustic_stack

The first term right of the "=" is equivalent to asking, "what uncertainty is there in running the Caustic technique over an ensemble that has a known mass of $M_{\text{ensemble}} = M_{\text{true}}$?" The next term asks, "What uncertainty is there in picking an M_{ensemble}?" Lastly, the third term asks, "What uncertainty is introduced by stacking clusters that may or may not be similar in mass?" Recall that this was one of the fundamental assumptions of the stacking technique: that the galaxies from each cluster are being drawn from a similar underlying dark matter distribution. This last term will be crucial in understanding if we can theoretically use the stacking technique on real data.

Our systematic analysis of the stacking technique addresses each of the terms on the right of Equation 9 one-by-one, so as to isolate each term's contribution to the overall uncertainty of the technique. This is done in three steps: 1.) Self Stacking, 2.) Mass Stacking, 3.) Observable Stacking.

We use two standard measures to gauge how well the Caustic technique performs. A statistical **bias** refers to accuracy: how close is your result to the true answer. A statistical **scatter** refers to precision: how well can you reproduce your work and get the same result. Consider a group of archers firing arrows at a far-off target. In one scenario, all arrows land very close together but are far from the bulls-eye, meaning that the archers have high precision but low accuracy. In another scenario, all the arrows strikes are evenly distributed across the target, meaning that their average position is right on the bulls-eye but their ability to control where the arrow strikes at any one time is poor. From a statistical point of view, we would like to minimize both our bias ($0 \pm 5\%$) and scatter ($< 30\%$), which is equivalent to maximizing precision and accuracy.

3.3. Self Stacking

Self stacking is a routine where we take one individual cluster and stack it and only it N_{clus} times to produce an ensemble cluster. In effect, we are stacking it against itself, hence the name self-stacking. In this case, each stacked phase space is just a different realization of the exact same underlying galaxy cluster with the same three dimensional gravitational potential and mass profile. The mass bin of the ensemble cluster contains only one value and is therefore a delta function centered on the individual cluster's M_{true}. This means that $P(M_{\text{ensemble}}|M_{\text{observable}}) \equiv 1$. In addition, because we are working in a simulation where we know the "true" mass of a galaxy cluster (see subsection 2.4), then $P(M_{\text{observable}}|M_{\text{true}}) \equiv 1$. This means that for self stacking, Equation 9 is reduced to

$$P(M_{\text{caustic}}|M_{\text{true}}) = P(M_{\text{caustic}}|M_{\text{ensemble}}) \tag{10}$$

which qualitatively states that any bias or scatter that we recover through self stacking is a direct reflection of the bias and scatter inherent to the technique, and is not a function of external galaxy cluster selection effects that plague us in the real universe and are encompassed by the other uncertainty terms.

Self stacking has one crucial limitation, though, and that is that we aren't actually creating an ensemble cluster if we stack the *exact same* phase space on top of itself. If that were the case, we would be putting points directly on top of points in the ensemble phase space, which wouldn't allow us to better characterize the Caustic surface and would defeat the purpose of stacking altogether. In order to get around this, we need to take new projections from different vantage points every time we stack so that we see a different projection of the same galaxy cluster N_{clus} times. The self stacking workflow thus follows these exact steps:

I. Set constants for this run of the code, such as N_{gal} & N_{clus}

II. Choose one galaxy cluster in the simulation's 3D cartesian coordinates

1. Load in that cluster's dark matter halo data
2. Initialize ensemble phase space data structures
   (a) Move the observer to a random position 50 h^{-1} Mpc away from the 3D cluster center
   (b) Project each galaxy in the field of view into a cluster-centric radius and line-of-sight velocity phase space
   (c) Choose N_{gal} galaxies from the phase space and append them to the ensemble's phase space
3. Repeat the previous step (step 2) N_{clus} times, which includes the first iteration
4. Remove interlopers in the final ensemble phase space via some interloper treatment technique
5. Run the Caustic technique on the final ensemble phase space, write the result out to file

III. Repeat previous step (step II) for all 2000 dark matter halos

The first thing we need to address is how to choose N_{gal} galaxies from a phase space in step 2 (c). In the simulations we have *all* of the galaxies that form around the galaxy cluster, including the really dim ones that we would probably not detect around a real galaxy cluster with a realistic telescope. A realistic method for choosing galaxies would be to take the top N_{gal} brightest galaxies, because these would be the first ones that we would observe in the real universe. This, however, does not work well in a self stacked ensemble cluster because each individual cluster phase space would then contain the same physical galaxies, just in different r-v positions. It turns out that if you stack the same physical galaxies with different r-v phase space positions you create artificial structure in the phase space that greatly biases the Caustic mass.

Figure 7 shows that when we always select the top brightest galaxies we get artificial structure in the phase space due to self stacking (Left-Top & Right-Top). We can circumvent this by taking a random selection of galaxies rather than the top brightest, however, this is dependent on our total sample of galaxies being sizably larger than N_{gal} itself. If the total sample size is close to N_{gal} then taking a random selection does not get rid of the artificial structure (Left-Bottom). So long as the total sample we randomly select from is greater than $\sim 10 \times N_{gal}$ we can produce a clean, structure-less phase space (Right-Bottom). All of the clusters in our 2000 galaxy cluster sample satisfy this criterion.

Before we run the Caustic technique over all 2000 galaxy clusters across the parameter space in our ensemble grid (see Figure 6), we need to deal with interloper galaxies. As a reminder,

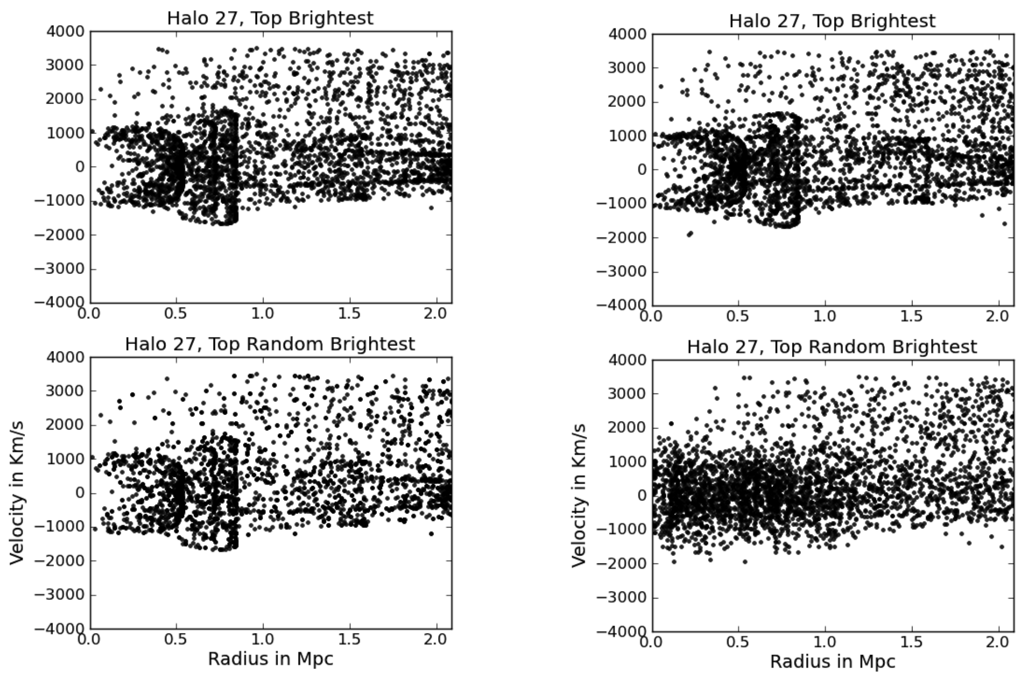


Fig. 7.— **Left-Top:** For each cluster projection, select a sample of $\mathrm{N_{gal}} \cdot 1$ brightest galaxies and from that sample, select the $\mathrm{N_{gal}}$ *top brightest* to stack. Clearly, artificial structure is apparent.
Left-Bottom: For each cluster projection, select a sample of $\mathrm{N_{gal}} \cdot 1$ brightest galaxies and from that sample, *randomly select* $\mathrm{N_{gal}}$ to stack. Artificial structure is still apparent when total sample = $\mathrm{N_{gal}} \cdot 1$.
Right-Top: For each cluster projection, select a sample of $\mathrm{N_{gal}} \cdot 25$ brightest galaxies and from that sample, select the $\mathrm{N_{gal}}$ *top brightest* to stack. Clearly, artificial structure is apparent.
Right-Bottom: For each cluster projection, select a sample of $\mathrm{N_{gal}} \cdot 25$ brightest galaxies and from that sample, *randomly select* $\mathrm{N_{gal}}$ galaxies to stack. Artificial structure goes away when total sample is $\mathrm{N_{gal}} \cdot 25$.

interloper galaxies are galaxies that are not actually apart of the galaxy cluster of interest, but are foreground (between observer and galaxy cluster) or background (behind the galaxy cluster) galaxies that are projected into the line of sight. They are typically easy to identify because, although they seem to have similar positions as the member galaxies, they typically have much different peculiar velocities–i.e. redshifts–and colors than member galaxies. They are, however, a few interloper galaxies that lie close enough to the cluster that they contaminate the escape velocity surface we are trying to measure. Stacking only compounds this issue, and actually makes it harder for a shiftgapper technique to eliminate interloper galaxies from near the escape velocity surface. In testing our technique, we found that a shiftgapper interloper-removal technique is unable to remove a substantial amount of interlopers from heavily stacked ensembles, which causes our mass estimates to be biased. To fix this, we employ a more rigorous interloper removal technique called "edge clip" (Gifford et al. in prep 2015). This technique takes the kernel density map of the phase space that the Caustic technique generates and takes its derivative. It then looks for the inflection point of the derivative, which effectively traces out the edge of the phase space along the velocity (vertical) axis, where galaxies go quickly from being populous to devoid. Depending on the slope of the derivative, it eliminates galaxies above a threshold near the determined inflection point.

Lastly, before we run the Caustic technique, we need to decide variation of the mass calculation we will use. As noted briefly in subsection 2.3, we currently have two ways to pick a set of iso-density contour on the phase space and correlate one of them to the escape velocity profile. The traditional method finds the iso-density contour that most closely matches the velocity dispersion of the cluster galaxies, which is what should be used for non-stacked galaxy clusters and is what we call a Caustic Surface. We quickly realized that this could not be used on stacked galaxy clusters because the velocity dispersion of an ensemble is ill-defined. Becker et al. (2007) analyzed stacked galaxy clusters in the maxBCG cluster catalogue from the SDSS data release 7, and found that their pair-wise velocity dispersions exhibit narrower-than-gaussian behavior. Analytically this makes sense. Each galaxy cluster has on average a gaussian distribution of galaxy velocities centered roughy on 0 km/s. When you add gaussians on top of each other–which is essentially what stacking is doing–the resultant distribution is known to be non-gaussian, and in fact, it is always narrower than gaussian. This is problematic for stacking because the halo velocity dispersion is defined to be the standard deviation of a gaussian distribution of velocities. Therefore, in a stacked ensemble, the traditional halo velocity dispersion is unphysical and will be biased, which will propagate into the traditional Caustic mass because it is calibrated using the cluster's velocity dispersion.

This forced us to come up with a new way to choose the correct iso-density contour without relying on the cluster's velocity dispersion. Gifford et al. in prep. (2015) found that one could reproduce an unbiased estimate of the escape velocity profile by binning the phase space in radial bins and, for each bin, taking the top X% of galaxies along the velocity axis, where X is generally $\sim 10\%$. The iso-density contour that minimizes the χ^2 along these points is then chosen as the "correct" iso-density contour and is correlated to the escape velocity profile. For Caustic mass estimates of ensemble clusters, we will always use masses derived using this Edge Surface. For Caustic mass estimates of individual clusters, we will always use masses derived from the traditional Caustic Surface calibrated to a halo velocity dispersion. We use the term Caustic mass interchangeably, but will explicitly clarify when their contexts overlap. Otherwise, it is safe to assume that ensemble cluster Caustic masses come from an Edge Surface, while individual cluster Caustic masses come from a Caustic Surface.

We run our stacking code over the ensemble grid across different $\mathrm{N_{gal}}$ & $\mathrm{N_{clus}}$ pairs and measure how the Caustic mass of an ensemble varies from it's true mass. Figure 8 characterizes the bias and scatter of the Caustic mass across the ensemble grid parameter space. In these plots, we are comparing the derived Caustic mass of each ensemble with the average mass of the individual clusters that were stacked, where the average is a robust median. The bias plot (left) shows that a large portion of the parameter space is close to $0 \pm 5\%$ bias. No matter how small the $\mathrm{N_{gal}}$ is, as long we we stack upwards of 25 clusters we can get to $0 \pm 5\%$ bias. We can therefore conclude that the process of stacking does not introduce any major uncertainties that severely detriment the accuracy of the Caustic technique. The overall behavior of the two plots show that the sampling size of the phase space, denoted by diagonal black lines (left plot) and horizontal axis (right plot), is strongly correlated to the overall bias and scatter of the Caustic technique, as was also determined in Gifford et al. (2013).

Figure 8-right tells us that the Caustic technique actually has lower scatter–higher precision–on ensembles than on individual clusters. The dotted black line is the same line from Figure 4 run

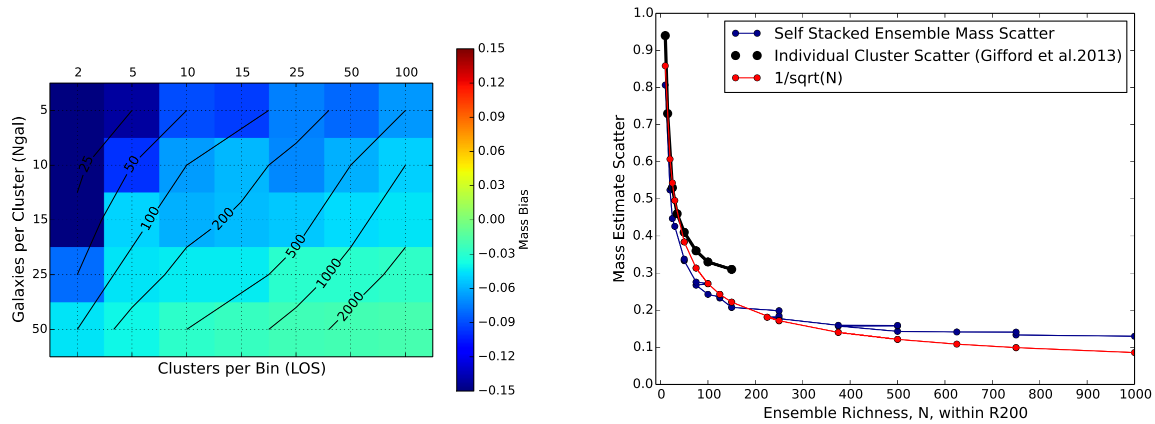


Fig. 8.— **Left:** Self stacking run over our ensemble grid parameter space, showing that we attain roughly 0% bias at sampling rates of > 1000 galaxies per phase space. **Right:** Scatter of ensemble masses as a function of galaxies in a phase space. The take-away is that the stacking technique has a much lower plateau than the original Caustic technique, mainly due to the averaging out of spherical asymmetries. This makes the stacking technique much more *precise* at high sampling.

over individual clusters, whereas the dotted blue line is the scatter on ensemble systems. On an ensemble, the Caustic technique levels out at around 15%, which is a factor of two improvement over its precision on individual clusters. The improvement in precision is due an ensemble's spherical symmetry; galaxy clusters are inherently aspherical and when they are stacked in random orientations their mutual asymmetries tend to average each other out. Because the Caustic technique assumers spherical symmetry, it does a more precise job on ensembles than on individual systems.

3.4. Mass Stacking

Mass stacking is one step closer to how one would construct an ensemble in the real universe. In the mass stacking routine, rather than stack the same cluster on itself, we stack together different clusters, each with a different mass. This means that we can take the $\mathrm{N_{gal}}$ brightest galaxies per cluster, which is an accurate description of how we would select galaxies in a real observation, without having to worry about creating artificial phase space structure as we did when self stacking.

This also means that we can no longer assume that the "true" mass of the ensemble is any one of the individual clusters that make it up, but some average of all of their masses. This correlates to saying that $\mathrm{P(M_{ensemble}|M_{observable})} \neq 1$, as we assumed with self stacking. $\mathrm{P(M_{observable}|M_{true})}$ still equals 1, however, because there is zero uncertainty related to what the true mass of each individual cluster is: our observable for each individual cluster is still the true simulation-quoted $\mathrm{M_{200}}$. The uncertainty equation (Equation 9) now looks like

$$P(M_{\text{caustic}}|M_{\text{true}}) = P(M_{\text{caustic}}|M_{\text{ensemble}}) \cdot P(M_{\text{ensemble}}|M_{\text{observable}}) \tag{11}$$

where $\mathrm{P(M_{caustic}|M_{ensemble})}$ has already been constrained in our self stacking analysis.

Figure 9 characterizes the bias and scatter of the Caustic technique when mass stacking. Similar to Figure 8 that shows increasing bias and scatter along the upper-left to lower-right diagonal, we find that mass stacked ensembles retain low bias at sampling rates greater than 500 galaxies per phase space. The mass scatter is similar to the scatter when self stacking and stays remarkably low even at medium sampling rates of 200 – 500 galaxies per phase space.

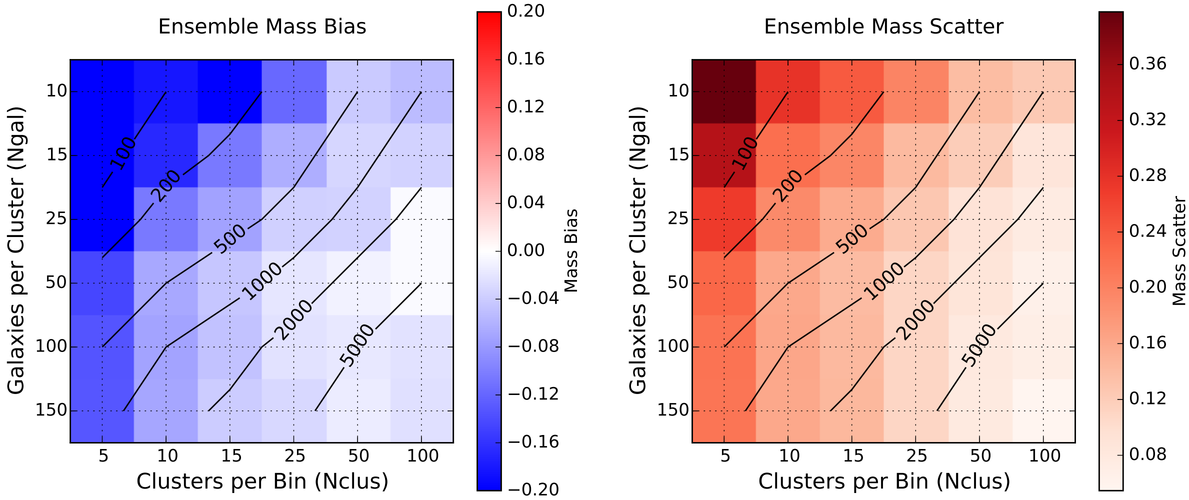


Fig. 9.— **Left:** Caustic mass bias with respect to the median of the individual cluster masses that make up each ensemble. **Right:** Caustic mass scatter about the median of the individual cluster masses that make up each ensemble.

The size of a bin's range in mass depends on how many high-mass and low-mass clusters are inherently formed in the universe and the number of clusters we choose to bin per ensemble. We could make a stacked ensemble in two ways: 1. maintain a constant number of clusters per ensemble and have variable bin size in mass or 2. maintain a constant bin size in mass and have a variable number of clusters stacked per ensemble. Above all else, we want to maintain the same total number of galaxies in the ensemble phase space because this sampling rate is very closely tied to our accuracy and precision. We therefore choose path (1.) and keep the number of clusters per bin a constant, while each bin has a varying range in mass. Because of this, high mass ensembles will have much larger mass ranges than low mass ensembles. For example, the ten most massive galaxy clusters in our sample span 2.3×10^{15} — 8.7×10^{14} $h^{-1}M_{\odot}$, while ten intermediate/low mass galaxy clusters in our sample span 8.676×10^{13} — 8.665×10^{13} $h^{-1}M_{\odot}$.

This could theoretically induce a non-constant bias, that is, a bias that changes as a function of mass. We find, however, that the difference in a bin's mass range does not induce a noticeable bias as a function mass. Figure 10 graphs the Caustic masses of 42 ensemble clusters against the median of the individual cluster masses that make up each ensemble. The blue line plotted is *not a best fit line*, but a one-to-one line, showing a theoretical 0% bias. This shows that when mass stacking ensembles with N_{gal}=15 & N_{clus}=50, one can recover unbiased and highly precise Caustic masses. The errorbars are calculated via a bootstrap resampling analysis, where we re-run the

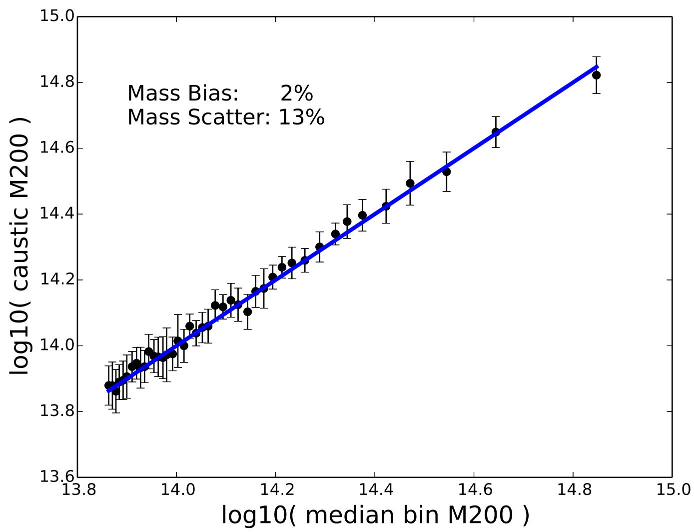


Fig. 10.— One-to-one graph for mass stacked ensemble clusters, where N_{gal}=15 and N_{clus}=50. Errorbars are calculated via a bootstrap resampling analysis. The blue line is not a best fit, but a one-to-one line showing theoretical 0% bias. This shows that the bias does not change as a function of mass.

technique many times and see how the results change each time. The standard deviation of these results is then plotted as the error bars.

3.5. Observable Stacking

We now incorporate the final term in Equation 9, $P(M_{observable}|M_{true})$. Recall that this term accounts for not knowing what the real masses of each galaxy cluster is in the real universe; how can we choose to stack clusters that are similar in mass if we don't know what their masses are beforehand? We need to turn to a mass proxy or observable that scales with mass. While we cannot get the absolute mass of a cluster with a mass-observable scaling law, we can estimate if one cluster is more or less massive relative to another, and in this way can sort through a catalogue and group clusters of similar mass. There will be an uncertainty associated with this scaling law, though, which is encompassed by this last term. This phenomenon is known as mass mixing (Becker et al. 2007), because we are sometimes mixing clusters of different masses into bins where they don't theoretically belong. This mass mixing effect is stronger the higher the scatter is on the mass–observable relationship. The goal of this analysis is to see how much bias and scatter mass mixing induces into the technique. If it induces a significant amount of bias and scatter that is both unpredictable and non-robust to our ensemble grid parameter space, we may not be able to stack clusters in the real universe.

Figure 11 shows our Caustic technique applied to observable-stacked ensembles in the MS. We use a richness estimator for our mock observable, described in subsection 4.3. These plots help us narrow in on which N_{gal} & N_{clus} configuration returns Caustic masses with low bias and scatter with respect to the average mass of the ensemble's constituent galaxy clusters. Our ensemble grid plots show that there is a somewhat significant bias induced by mass mixing (Figure 11-Left). However,

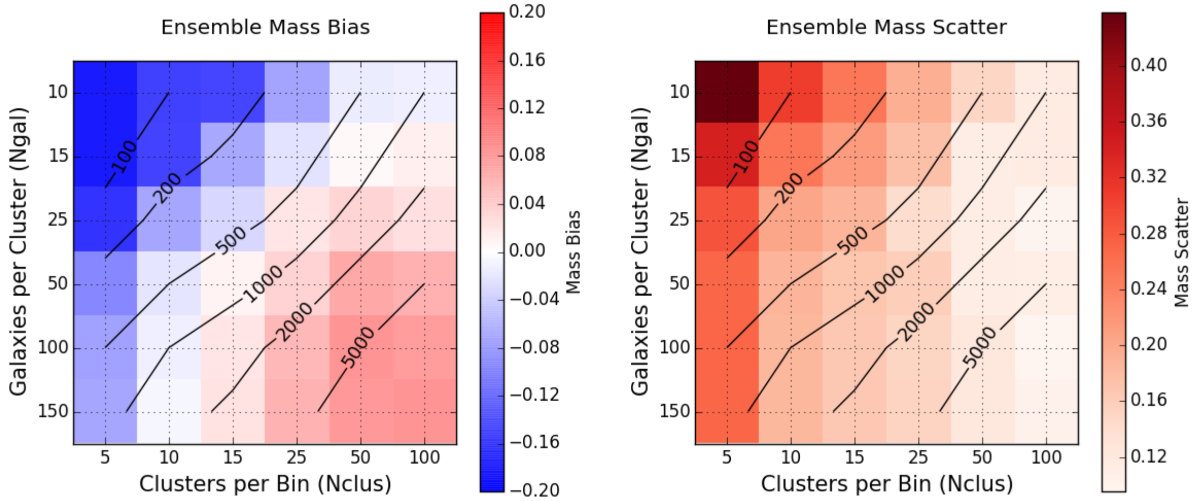


Fig. 11.— **Left:** Mass bias of ensemble grid with mass mixing induced. Ensemble Caustic masses are biased in certain regions of the parameter space, but are unbiased in other. **Right:** Mass scatter over ensemble grid parameter space. Ensemble Caustic mass precision remains relatively robust to mass mixing. Diagonal lines show the total phase space sampling size, $N_{gal} \cdot N_{clus}$

this bias really only starts to have a significant effect when the ensemble sampling is greater than 1000 galaxies per phase space. With an N_{gal}=15 and N_{clus}=25 or 50, we can still get relatively unbiased results with a scatter near 15%.

The reason for bias at a high N_{clus} is likely due to an increase in the mass ranges of the ensembles. When stacking a larger number of clusters, there is a higher chance that we accidentally stack a high mass cluster into a low mass bin. This is just the nature of mass mixing. This could adversely affect the Caustic mass estimate of this ensemble because the high mass cluster populates more galaxies above the phase-space edge of the smaller galaxies, forcing the Caustic technique to overestimate the escape velocity profile and derive a mass that is biased high.

However, the important part of Figure 11 is not that there is a new bias induced by mass mixing, its that the bias is still relatively robust to how we construct our ensembles. In other words, if you move one or two grid cells from the center of Figure 11, you only induce a bias change of roughly 10% at worst. What this means is that while there might be a slight bias that is induced from the mass mixing effect, 1. it isn't significant enough (ex. 20+%) to prevent us from using the stacking technique, 2. it has predictable behavior, in that it increases as the total sampling of the phase space increases, and most importantly, 3. the bias per ensemble grid point is independent of mass. This last point, #3, is explained by Figure 12, which takes one grid point in our ensemble parameter space and plots its 40+ ensemble Caustic masses. The blue line shows theoretical 0% bias. What is clear is that the ensembles are biased high, which means it was taken from a pink-ish grid point in our parameter space of Figure 11. However, it is evident that the slope of Figure 12 has the same slope as the blue line. This means that while the ensembles are indeed biased high, that bias is independent of mass and is a constant throughout our sample of ensemble clusters. This means that if we can calibrate one portion of these ensembles, say the high mass end, then we can calibrate the entire sample of ensembles because the bias is the same throughout the sample.

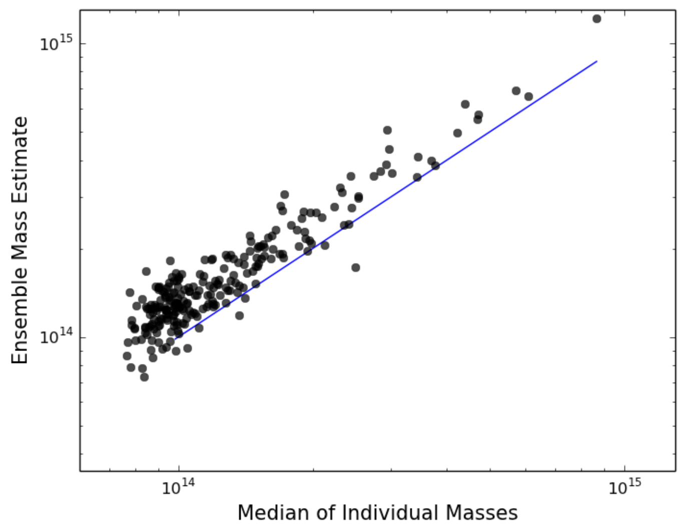


Fig. 12.— True mass versus stacked ensemble Caustic masses for one N_{gal} & N_{clus} configuration of the ensemble grid parameter space (Figure 11) showing that while the ensemble Caustic masses may be biased, that bias is the same across all mass ranges, i.e. the bias is mass independent, a prediction we will look to confirm when we stack on real galaxy survey data.

3.6. Conclusion

The Caustic technique's ability to recover accurate and precise Caustic mass estimates is highly dependent of sufficient phase space galaxy sampling (Figure 3). For clusters that have less than 50 known member galaxies, the Caustic technique is rendered effectively unusable because its bias and scatter depreciate significantly. Unfortunately, low mass clusters and all clusters at high redshift suffer from sparse sampling effects. If we want to use dynamical mass estimators on these clusters, we need to find a way to circumvent the Caustic technique's weakness to sparse sampling.

In order to remedy this problem we developed a model for a stacking routine that can be used on galaxy cluster dynamical mass estimators, specifically the Caustic technique. By combining the galaxy data from multiple sparsely sampled galaxy clusters into one ensemble cluster, we can significantly increase the precision and accuracy of the Caustic technique. For example, for a set of 25 galaxy clusters where we have spectroscopic redshifts for only 15 member galaxies, their individual mass estimates would suffer from a bias of $> 60\%$ and a scatter of $> 70\%$, according to Gifford et al. (2013). Assuming we can get a rough estimate of their mass via a mass proxy like their N_{200}, we can stack these clusters into an ensemble galaxy cluster and run the Caustic technique to get an estimate of their average mass, which has a theoretical 0% $\pm$5% bias and 15% scatter (Figure 11).

We came to this conclusion through a highly systematic and rigorous test of our technique in three main phases: self stacking, mass stacking and observable stacking. Each phase constrains different uncertainty aspects that effect the entire stacking process (Equation 9). In doing so, we have come to a deep understanding of how the process of stacking a galaxy cluster in the real universe can bias our technique. By exploring the parameter space of our ensemble grid (Figure 6, Figure 11) we find that we need to operate in certain N_{gal} & N_{clus} regimes when stacking real data in order to stay unbiased. Furthermore, the bias that is induced from mass mixing is independent of the mass of the ensemble cluster. This means that if we can get a small subsample of highly accurate galaxy cluster masses at the high mass end, we can calibrate our ensemble One-To-One line (Figure 12) and therefore effectively calibrate the entire sample of ensembles.

Other authors have stacked galaxy clusters and applied the Caustic Technique to learn about the dynamics of stacked galaxy clusters (Serra et al. 2011; Rines et al. 2013). They stack galaxies in order to deduce the ensemble's mass profile to high precision, which needs a high sampling rate of galaxies to achieve. They do not, however, perform any kind of systematic test of their stacks to understand how the process of stacking affects their ensemble mass estimates. This work is the first work that has taken a detailed and rigorous look at how the process of stacking affects dynamical mass estimates of ensemble galaxy clusters. We have shown that the way in which we build our ensemble cluster does indeed affect the bias and scatter of ensemble mass estimates and must not be overlooked in future studies of ensemble galaxy clusters. In conclusion, we find that stacking algorithms have a high potential to improve our ability to constrain the masses of clusters across different mass and redshift ranges, given that we first understand the inherent uncertainties associated with stacking.

4. Application to Optical Galaxy Surveys

4.1. Introduction

Over the past decade, galaxy cluster surveys have brought in incredible amounts of data on the masses of clusters in the radio, optical and X-ray wavelength regimes. In the present and near future, large galaxy cluster surveys will continue to push our datasets on galaxy clusters to higher redshift and lower masses. Here, we present an initial application of our stacking technique on real data, showing that it is a useful tool in recovering the masses of sparsely sampled galaxy clusters in the real universe.

4.2. SDSS DR12 Data Set and the C4 Cluster Catalogue

The Sloan Digital Sky Survey (SDSS) is a photometric imaging and spectroscopic observational program that has mapped a large part of the sky and taken an unprecedented number of spectroscopic galaxy redshifts. Beginning in 2000 and continuing up to the present, SDSS has mapped over 500 million objects; their most recent data release, DR12, was made public in July 2014 and contains over 4 million spectra. We use SDSS's DR12 dataset to create ensemble galaxy clusters at low redshift to test our stacking algorithm on a real dataset. We subsample from the entire DR12 dataset and limit ourselves to a magnitude and redshift cut, where each galaxy's SDSS R band absolute magnitude is $<$ -19.1 and spectroscopic redshift is $<$ 0.25.

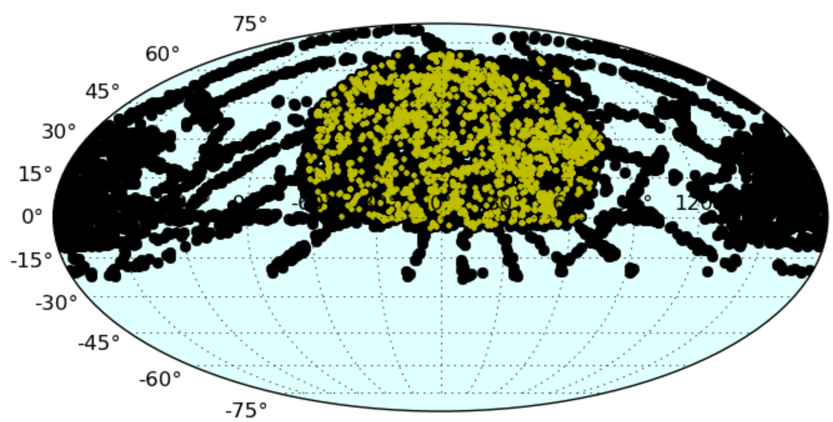


Fig. 13.— Sky coverage of the C4 cluster catalogue. Black points are photometry from SDSS DR12, yellow points are C4 identified galaxy clusters.

Before we stack clusters, though, we need to know where the clusters are. In order to locate clusters in the universe we rely on a cluster finder algorithm. For our study, we use the C4 cluster finder (Miller et al. 2005). The C4 cluster finder identifies galaxy clusters in a multi-dimensional position, redshift and color space. It is a red sequence based algorithm, meaning that is uses the fact that galaxies in the cores of clusters tend to be red to identify galaxy cluster member galaxies. Miller et al. (2005) ran the C4 cluster finder algorithm over SDSS data and created the C4 cluster catalogue, which is simply a list of galaxy clusters and their positions on the sky. The C4 cluster catalogue is a low redshift catalogue, extending from a z of ~ 0.02 to ~ 0.17, and contains ~ 2500 galaxy clusters. Figure 13 shows our C4-identified galaxy clusters in the sky on top of SDSS DR12 photometric data. Some of these clusters are not actually clusters, though, and are either substructure effects or chance projections of overlapping clumps of galaxies on the sky. After setting basic constraints on the cleanliness of the sample, we are left with a sample of 1500 galaxy clusters.

4.3. Richness Estimator

In order to stack galaxy clusters together, we need to first use a mass-observable scaling relationship to get a rough estimate of their mass so that we can arrange our sample into sets of galaxy clusters of similar masses. We can do this many different ways, with a cluster's total luminosity, galaxy richness or galaxy velocity dispersion. For our analysis, we create a galaxy richness estimator and use each individual galaxy cluster's N_{200} to determine their masses relative to other galaxy clusters.

In essence, a galaxy cluster's richness is simply a measure of how many galaxies are in the cluster. For high mass clusters with larger gravitational potential wells, we would expect more galaxies to be trapped under its gravitational influence. For low mass clusters, we would expect less. A galaxy cluster's richness, therefore, scales with its overall mass. Within the line of sight toward any one galaxy cluster lies thousands of foreground and background galaxies that are completely unrelated to the galaxy cluster of interest, which induces a scatter about this idealistic relationship.

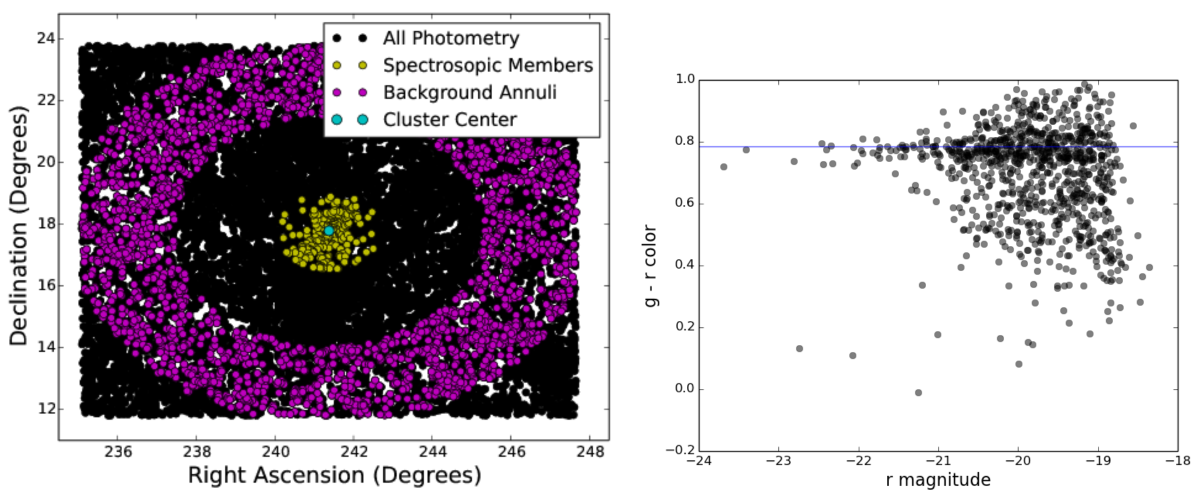


Fig. 14.— SDSS DR12 photometric data for one galaxy cluster (cyan) surrounded by member, foreground and background galaxies (colored points). A cluster's richness is all red galaxies within 1 virial radius (yellow) minus the density of red galaxies in a annuli from 3 to 5 virial radii (magenta).

Galaxies within a virial radius of a galaxy cluster tend to be "red and dead." This relationship is called the Red Sequence of galaxy clusters and we can pick out clusters via an apparent overdensity in galaxies' color-magnitude space. To measure a cluster's richness, we simply calculate the number of galaxies that have a similar color as the cluster itself that lie within the cluster's r_{200}. To account for foreground and background galaxy contamination, we sweep out an annuli of galaxies that lie within $3 \times r_{200} < r < 5 \times r_{200}$, find the density of galaxies that have the same color as the cluster of interest, and subtract that density times the projected area of the cluster from the initial richness measurement. Once we have a measure of N_{200} for each individual galaxy cluster, we re-sort the ordering of the cluster catalogue based on their N_{200} so that we can relate clusters of similar mass.

To determine the color of a cluster we use the same radius-velocity phase space the Caustic Technique uses, and take all galaxies with a radius within r_{200} and velocity within two times the velocity dispersion of the cluster. We then take a robust mean of the those galaxies' color: color = SDSS G Filter Magnitude - SDSS R Filter Magnitude. Figure 14–Right shows an example color–magnitude plot, where the cluster color is denoted by the horizontal blue line. We then go back to the photometric dataset and pick out all projected galaxies within 1 virial radius and within 1σ of the cluster color (Figure 14–yellow points). The number of galaxies that fit this category is the "signal" of the cluster. We then take all galaxies within a $3 \times r_{200} < r < 5 \times r_{200}$ annuli that are also within 1 sigma of the cluster color (magenta points), divide that number by the area swept out by the annuli and call this the background density. The richness is then the background density times the area swept out by 1 virial radius, subtracted from the initial signal: richness = signal - background.

4.4. Stacking on 900 Low-z C4 Clusters

Using our richness estimator and the Caustic Technique, we calculate the richnesses and Caustic masses for 900 individual galaxy clusters from the C4 Cluster Catalogue using the SDSS DR12 spectroscopic sample extending from $0.03 < z < 0.15$. (Figure 16: Top-Left). By eye, it seems evident that there is some non-linear behavior, or at least, bi-modal linear behavior of two linear functions that pivot at around $\log_{10}(N_{200}) = 1.75$. To show this more distinctly, we separate the figure into two cluster populations: those that have over 50 galaxy spectra and those that have less than 50 galaxy spectra. We use a robust line fitter and plot best fit lines to each population to highlight the non-linearity of the sample as a whole (Figure 16: Top-Right). While it may not be statistically correct to fit lines to these populations, we do so to merely highlight the fact that these populations likely do not belong to the same underlying population. However, drawing insight from our systematic test of the Caustic Technique in §2, for the population with over 50 galaxies per cluster (blue) we would expect it to follow linear behavior because we have achieved the sampling necessary to recover theoretically unbiased masses. For the population with less than 50 galaxies per cluster (black) we are in the regime where the Caustic Technique returns masses that are significantly biased low and would expect the non-linear behavior that matches the bias(N_{200}) behavior from Figure 3. Both of these expectations assume that the underlying "true" scaling of N_{200} with M_{200} is indeed linear in log space (i.e. described by a single power law in real space), which we find that it is in the Millennium Simulation (see Figure 15).

The fact that we recover non-linear behavior at the low richness end of our Mass–Richness relationship is therefore indicative of the Caustic Technique's bias at low richness. To test how our stacking algorithm works on this data set, we bin the 900 clusters by their measured N_{200} and create ensemble clusters each composed of 25 clusters and 15 galaxies taken per cluster. We measure their

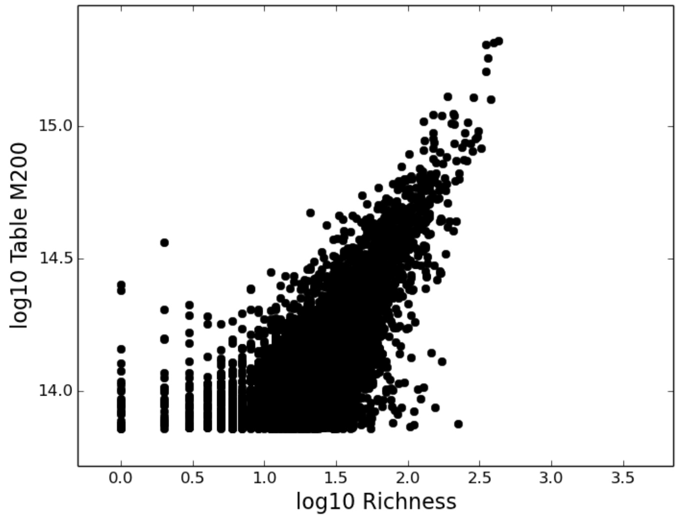


Fig. 15.— Mass–Richness relationship in the Millennium Simulation showing an underlying linear behavior in log-log space.

Caustic "Edge" masses and plot them against the median N_{200} of each bin's constituent clusters (Figure 16: Bottom-Left). We fit a line to those points and recover a linear relationship in log space. Overlaying all of these points and fits on top of each other, we see that the ensemble best fit (cyan) agrees with the fit to the unbiased Caustic masses (orange) to first order (Figure 16: Bottom-Right). It could be argued that its normalization is slightly lower, which would correspond to it being slightly underbiased. However, as we mentioned in subsection 3.5, the exact bias of the ensembles with respect to the individual clusters' Caustic mass hovers around $0\% \pm 5-10\%$. What is important is the fact that the cyan line has similar slope as the orange line, which is a statement that the bias, if present, is independent of mass (or richness), which is exactly as we predicted with our simulated tests. We fit lines to the magenta and blue points using a robust linear regression technique called Random Sample Consensus (RANSAC), which is an iterative method that accounts for outliers. Future work will employ a fully Bayesian linear regression model similar to Andreon & Hurn (2010).

One of the benefits of stacking is that we can attain unbiased mass estimates for clusters that have only a handful of measured galaxies. As we showed in section 3, these ensemble mass estimates are theoretically an unbiased tracer of the average masses of the individual galaxy clusters that make up the ensemble. We can therefore take a mass-observable relationship, which does not necessarily give absolute masses and could be biased, and calibrate it so that it is unbiased. We would do this by creating a transfer function for the biased part of Mass–Richness relationship (i.e. black points in Figure 16). This transfer function is the multiplicative difference between a linear model derived from the ensembles, and a non-linear model derived from the individual clusters that are biased:

$$M_{200,\mathrm{unbiased}}(\mathrm{N}_{200}) = \mathcal{T}(\mathrm{N}_{200}) \cdot M_{200,\mathrm{biased}}(\mathrm{N}_{200}) \tag{12}$$

where $\mathcal{T}$ is the transfer function. In other words, the transfer function is the difference between

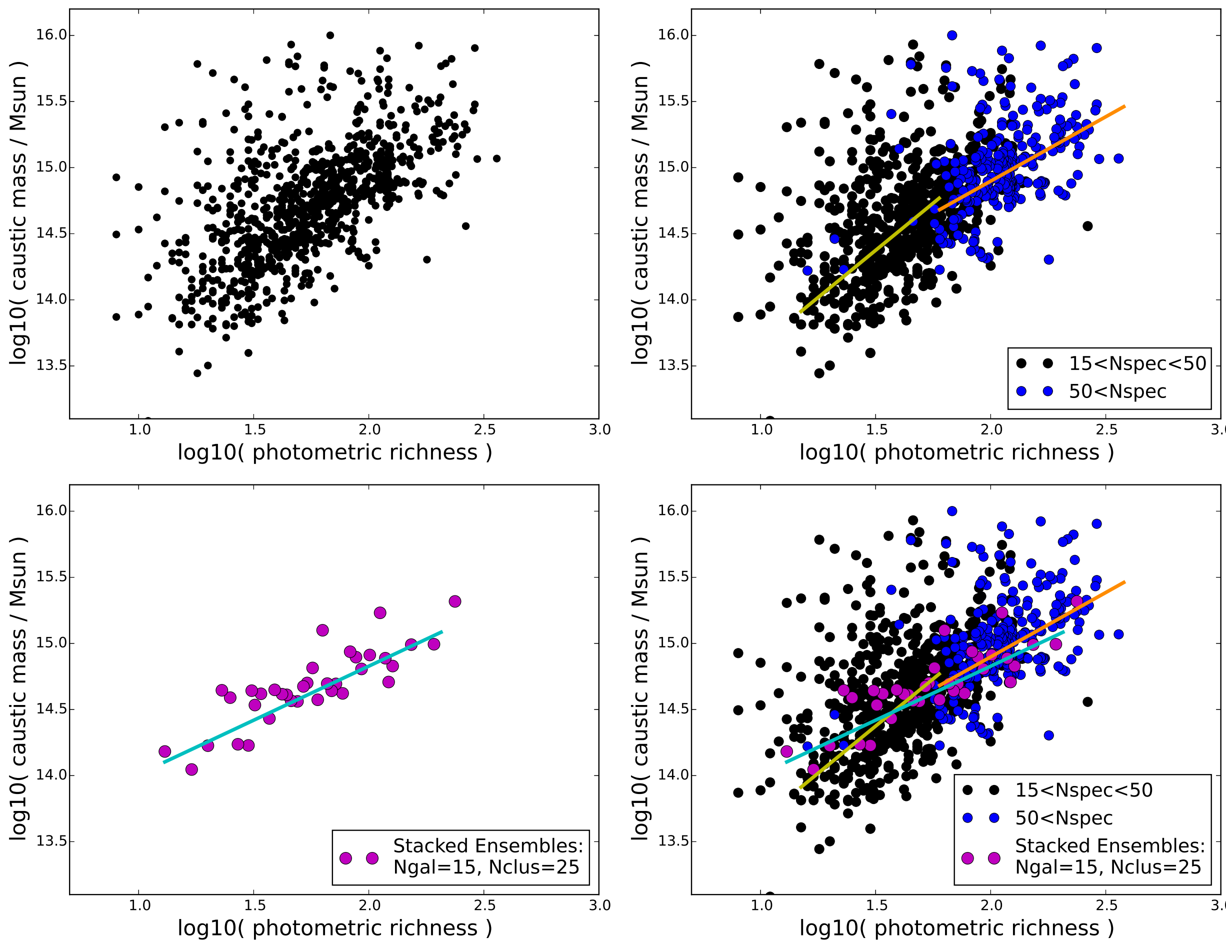


Fig. 16.— **Top-Left:** Mass–Richness relationship for 900 low-redshift C4 galaxy clusters. As a whole, the graph seems to show some non-linear behavior. **Top-Right:** We split the same graph into two populations: those that have over 50 spectroscopic redshifts and those that don't. We use a robust line fitter to get a rough idea of how these populations scale with N_{200}. **Bottom-Left:** We then stack over these 900 clusters binned by their measured N_{200}. **Bottom-Right:** Two previous plots overlaid, showing that the ensemble best fit line (cyan) roughly matches the blue population best fit line (orange) in terms of both slope and normalization.

the black points and the magenta points at the lower-right corner of Figure 16.

Once we have a Mass–Richness relationship that is unbiased with respect to galaxy cluster sampling, we can apply this scaling law to all galaxy clusters in the C4 sample and attain a statistically unbiased set of masses for our initial sample of 1,500 galaxy clusters. With an unbiased set of galaxy cluster masses we can theoretically begin to do galaxy cluster cosmology, such as counting the number of galaxies in a given mass range, called the Mass Function (Figure 17). This kind of cosmology is very sensitive to time evolution: from low to mid redshift, the Mass Function changes significantly depending on cosmological parameters (Tinker et al. 2008). However, the C4 sample alone does not provide enough galaxy clusters at high redshift to really perform any cosmological tests. To really test cosmology models, then, we will need to push out to higher redshift and apply the stacking algorithm to a cluster catalogue that extends to a redshift of around 0.7.

The RedMapper catalogue provides us with just that ability.

4.5. Conclusion & Near-Future Applications

In the near future, we will extend the stacking code to run on higher redshift data sets. The RedMapper cluster finder, for example, identifies galaxy clusters similar to the C4 method by using the red sequence, but picks clusters from a redshift of 0.15 to 0.7 (Rykoff et al. 2014). These galaxy clusters have SDSS DR12 spectroscopy but most are poorly sampled, making them ideal candidates for our stacking code. With a data set that extends out to mid redshifts of ~ 0.7, we can begin to perform real cosmological tests.

The mass function describes the number of galaxy clusters produced at a given mass. This is highly sensitive to cosmological parameters, such as the relative energy densities of dark energy and dark matter. Figure 17 shows the mass function of galaxy clusters in the Millennium Simulation, where the circles represent the galaxy cluster number counts and the red and black lines represent theoretical predictions at z=0 and z=0.5 with the red showing no dark energy and the black line showing strong dark energy. This says that if we can get accurate number counts out to higher and higher redshifts, our stacking technique may be able to theoretically constrain predictions of dark energy, Λ. RedMapper is a great dataset to try and do this test on, and we now have the data, the tools and a theoretical understanding of the stacking technique needed to do it.

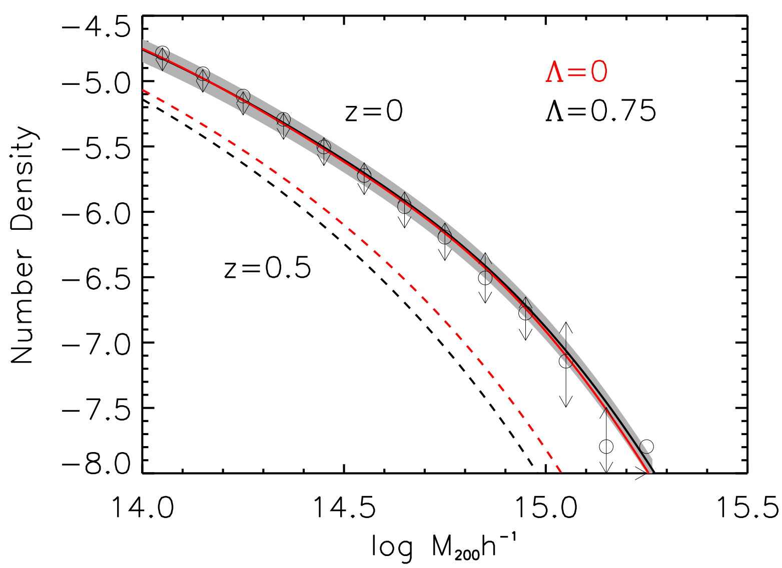


Fig. 17.— Millennium Simulation mass function (circles) and theoretical Tinker mass function (lines) at z=0 and z=0.5, where red is no dark energy and black is strong dark energy. Measuring the mass function of galaxy clusters at z=0.5 with the stacking technique may be able to constrain Λ, the dark energy cosmological constant. Figure from Miller et al. 2014.

5. Conclusion

Over the next two decades, large ground-based telescopes will deliver extremely vast amounts of astronomical data. A significant focus of this data will be geared towards understanding arguably two of the most pressing questions in astrophysics in the 21st century: what is dark matter and dark energy? To resolve these questions, the astronomical community is going to have to use a multi-faceted approach and use different kinds of cosmological probes to build a comprehensive and coherent picture. A key part of this approach is the study of galaxy clusters and the constraints they can put on cosmological parameters.

There are only a few ways to directly measure the dark matter mass of a galaxy cluster's dark matter halo. One way is the Caustic technique, which uses the positions and velocities of orbiting galaxies to trace the galaxy cluster's escape velocity profile, which is directly related to its gravitational potential and Newtonian mass. We showed that the Caustic technique can accurately (0% bias) and precisely (40% scatter) recover the M200 of a galaxy cluster given that we have spectra for over 50 of its member galaxies. Many of the known galaxies clusters in the universe, however, only have a handful of galaxy spectra associated with them. This means that we cannot feasibly use the Caustic technique, or any dynamical mass estimator for that matter, to derive their M200. This is unfortunate because precision cosmology depends heavily on high number statistics; we cannot do good cosmology with only a handful of data points. It is also unfortunate because galaxy clusters that tend to be poorly sampled are typically the ones that are the most interesting: the ones at the edges of the redshift limits of the current telescopes, were we will always be making new discoveries.

For these reasons we set out to build a stacking technique that would allow us to use the Caustic technique on galaxy clusters with only a handful of galaxy spectra. In this thesis, we detailed how we systematically built and tested our technique against N-body and semi-analytic datasets from one of the largest cosmological simulations ever run. We show that by stacking galaxy clusters together to form an ensemble cluster, we can theoretically increase our precision by a factor of three, and retain our accuracy so long as restrict ourselves to unbiased parameter spaces in Figure 11.

We then applied our theoretical framework on real data and showed that we observationally confirmed certain predictions from our simulations: namely that mass mixing induces minimal bias and that the bias is independent of mass. This means that the stacking technique can feasibly be used on real data. We then used the Caustic Technique to derive a Mass-Richness relationship for the C4 cluster catalogue (Figure 16). After stacking C4 clusters, we self-calibrated our Mass-Richness relationship at the low-richness end and produced a Mass-Richness relationship for 1,500 galaxy clusters in the C4 sample that is theoretically unbiased with respect to spectroscopic sampling effects. This is the first time such a Mass-Richness relationship has been produced via dynamical methods on real galaxy cluster survey data. In the near future, we will employ the same procedure over the RedMapper cluster catalogue, which extends out to a redshift of 0.7 and contains tens of thousands of galaxy clusters. If successful, we could theoretically use galaxy cluster masses to constrain cosmological models of dark energy.

More generally, we have presented for the first time a systematic and rigorous test of how dynamical mass estimators respond to stacked ensemble galaxy clusters. This has far-reaching implications for all cosmologists, in that we can now use the many spectroscopically-poor galaxy clusters that make up many of known galaxy clusters to do cosmology. This has the significant result of enabling cosmologists another tool in our scientific toolbox to study the nature of dark matter and dark energy, where each tool is crucial in constructing a comprehensive yet coherent picture of our universe.

REFERENCES

Andreon, S., & Hurn, M. A. 2010, MNRAS, 404, 1922 [4.4]

Becker, M. R., McKay, T. A., Koester, B., et al. 2007, ApJ, 669, 905 [2.5, 3.3, 3.5]

Bertone, S., De Lucia, G., & Thomas, P. A. 2007, MNRAS, 379, 1143 [2.4]

Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645 [2.4]

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2 [2.4]

Diaferio, A. 1999, MNRAS, 309, 610 [2.3, 2.3, 2.3, 2.3]

Diaferio, A., & Geller, M. J. 1997, ApJ, 481, 633 [2.3]

Evrard, A. E., Bialek, J., Busha, M., et al. 2008, ApJ, 672, 122 [2.3]

Gifford, D., Miller, C., & Kern, N. 2013, ApJ, 773, 116 [1, 2, 2.4, 3, 4, 3.3, 3.6]

Gifford, D., & Miller, C. J. 2013, ApJ, 768, L32 [2.3]

Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101 [2.4, 3.2.1]

Henriques, B. M. B., White, S. D. M., Lemson, G., et al. 2012, MNRAS, 421, 2904 [3.2.1]

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18 [1]

Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, ApJ, 650, 128 [2.3]

Miller, C. J., Nichol, R. C., Reichart, D., et al. 2005, AJ, 130, 968 [4.2]

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563 [2.2]

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A29 [2.3]

Rines, K., Geller, M. J., Diaferio, A., & Kurtz, M. J. 2013, ApJ, 767, 15 [3.6]

Rozo, E., Rykoff, E. S., Evrard, A., et al. 2009, ApJ, 699, 768 [2.3]

Ryden, B. 2003, Introduction to cosmology [2.2]

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104 [4.5]

Saro, A., Mohr, J. J., Bazin, G., & Dolag, K. 2013, ApJ, 772, 47 [2.3]

Serra, A. L., Diaferio, A., Murante, G., & Borgani, S. 2011, MNRAS, 412, 800 [2.3, 3.6]

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726 [2.4, 3.2.1]

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629 [2.4, 3.2.1]

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709 [1, 4.4]

Zwicky, F. 1937, ApJ, 86, 217 [1]

This preprint was prepared with the AAS LATEX macros v5.2.