Mon. Not. R. Astron. Soc. **000**, 1–13 (2015) Printed 22 June 2016 (MN LATEX style file v2.2)

Did we see self-interacting dark matter or statistical noise? The galaxy-dark matter offsets of the galaxy clusters in the Illustris simulation

Karen Y. Ng,[1] Annalisa P. Pillepich,[2] William A. Dawson,[3] D. Wittman,[1] Lars Hernquist,[2]

arXiv

ABSTRACT

Assuming that dark matter has a zero interacting cross section, how likely is it for us to see the offset values between dark matter and galaxies from real data? This paper formulates a test using cluster data in the cosmological simulation, the Illustris simulation, to examine that hypothesis. We examine the uncertainties of the different summary statistics, and see if the galaxy population have statistics consistent with those of the DM population. TODO: result summary. We found that the uncertainty of the offset resulting from projection effects are non-negligible and vary in unpredictable ways.

Key words: galaxy clusters, dark matter, statistics

1 INTRODUCTION

During the latest stage of structure formation, the universe gave birth to non-linear, hierarchical structures known as galaxy clusters. These clusters, made up of dark matter, galaxies and hot gas, are constantly accreting, merging and evolving with their environments. Bright galaxies that belong to a galaxy cluster or group, in particular, highlight the overdensities of the underlying dark matter (DM) distribution.

In these dense regions of the clusters, the rates of particle interactions can be enhanced, including the long-suspected self-interaction of DM particles (hereafter, SIDM). Many papers have used the offsets between the summary statistics of the DM density and the galaxy density to give constraints on the self-interaction cross section, i.e. $\sigma_{\rm SIDM}$, of dark matter. A lot of observational studies focus on using merging galaxy clusters as they assume the high collisional velocity should further increase the chance of detecting the effects of SIDM. By assuming galaxies being relatively collisionless $\sigma_{\rm gal} \approx 0$, any offset of the DM population from the galaxy provides $\sigma_{\rm SIDM}$ relative to $\sigma_{\rm gal}$. These observational studies include Markevitch et al. (2004) and Bradač et al. (2006) reporting an offset of 25 kpc for the Bullet Cluster; Dawson (2013) reporting an offset of 129 kpc and 47 kpc for the southern and the northern subcluster respectively; and Jee et al. (2015) reporting an offset of 190 kpc for MACSJ1752. However, other studies using 129 X-ray selected relaxed galaxy clusters, such as George et al. (2012) also report offsets of the same order of magnitude, between $50 - 150$ kpc.

On the other hand, there are many staged simulations of mergers of galaxy clusters that focused on detecting the signal from SIDM. They have consistently reported offset signals from SIDM simulations on par with statistical uncertainties. These simulations have raised questions about how strongly the galaxy-DM offsets can constrain the effects of SIDM. These staged simulation have parametric prescriptions of the spatial distribution of galaxies (Randall et al. 2008, Kahlhoefer et al. 2013, Markevitch et al. 2004, Robertson et al. 2016), such as an NFW profile, and may not have realistic DM substructures that surround the clusters. Common to these staged simulation is the negligible galaxy-DM offset when assuming $\sigma_{\rm SIDM} = 0\ {\rm cm}^2/\ {\rm g}$. Randall et al. (2008) found an offset of only 1.8 kpc in the staged simulation with $\sigma_{\rm SIDM} = 0\ {\rm cm}^2/\ {\rm g}$ using 10^5 galaxy particles that have identical distribution as the DM population at the beginning of their simulations. Kim and Peter et al. (2016) and Kahlhoefer et al. (2013) also show null offset throughout their staged simulation with $\sigma_{\rm SIDM} = 0\ {\rm cm}^2/\ {\rm g}$. It is likely that these idealistic simulation settings do not show the contribution of statistical and observational uncertainties to the galaxy-DM offsets. When Kahlhoefer et al. (2013) simulated SIDM with both low-momentum-transfer self-interaction and rare self-interactions of DM with high momentum transfer, they found maximum offsets that are < 30 kpc for $\sigma_{\rm SIDM}$ as high as 1.6 cm^2 / g. The reported offset from Randall et al. (2008) for $\sigma_{\rm SIDM} = 1.24\ {\rm cm}^2/\ {\rm g}$ is only 53.9 kpc. While [TODO] Kim and Peter et al. (2016) found a maximum offset < 50 kpc for $\sigma_{\rm SIDM} = 3\ {\rm cm}^2/\ {\rm g}$, and Robertson et al. (2016) also found a maximum offset $\lesssim 50$ kpc from a simulation suite of the Bullet Cluster with $\sigma_{\rm SIDM} = 1\ {\rm cm}^2/\ {\rm g}$.

It remains unclear how large the galaxy-DM offset should be without SIDM purely due to observational uncertainties and statistical noise. We therefore, perform mock observations of the galaxies clusters of the Illustris simulation to characterize the intrinsic scatter of the offsets without SIDM. Simply put, we perform a hypothesis test with the galaxy-DM offsets in the Illustris simulation directly

corresponding to our null hypothesis $\mathcal{H}_0$, with:

$$\begin{cases} \text{the null hypothesis } \mathcal{H}_0 : \text{Cold Dark Matter (CDM)} \\ \text{the alternative hypothesis } \mathcal{H}_1 : \text{Self-interacting Dark Matter (SIDM)} \end{cases} \tag{1}$$

This is the first study to investigate if it is a feasible idea to use the galaxy-DM offsets given the main constraints from observations, namely, the projections and the dynamical states of the galaxy clusters. These latent variables are confounding and can increase the variance of the population summary statistics of the galaxy-DM offsets of galaxy clusters.

This exercise is further complicated by the fact that there is no theoretical foundation showing which observable would be the most sensitive to SIDM. In fact, Kahlhoefer et al. (2013) have argued that SIDM does not cause significant galaxy-DM offsets. We explore the ways of computing the smallest galaxy-DM offsets from the Illustris simulation data.

A second utility of computing the galaxy-DM offset is to for computing the cluster-mass distribution using optical survey data. Understanding the closest optical proxy of the DM peak will enable mass-estimation via the stacking of weak lensing signal.

Popular choice for computing the offsets involves first inferring the summary statistic of each of the DM and the galaxy population of a cluster before taking a difference. While there are well established procedure driven by lensing physics for inferring the DM spatial distribution, there is no standard procedure for mapping the sparse member galaxy distribution. We quantify the bias and uncertainty associated with the one-point summary statistic for summarizing the physical state of a galaxy cluster.

In this paper, we 1) extract realistic observables from the Illustris simulation for comparison with observations, 2) explore the pros and cons of the different statistic for summarizing *the member galaxy population* of a galaxy cluster, 3) give estimates for the offsets between the summary statistics of the galaxy population and the DM population under ΛCDM cosmology, which we call

$$\Delta \boldsymbol{s} \equiv \boldsymbol{s}_{\rm gal} - \boldsymbol{s}_{\rm DM}. \tag{2}$$

where $\mathbf{s_{gal}}$ and $\mathbf{s_{DM}}$ are the two-dimensional (2D) spatial locations of the summary statistic of the galaxy population, and the density peak of DM respectively. This gives an estimate of the baseline scatter of offsets without any SIDM. And finally we 4) examine the properties of the clusters that give outliers in the offset distribution and 5) investigate the correlations between the 3-dimensional properties of a cluster and the projected observables such as Δs.

The organization of this paper is as follows: In section 2, we will describe the physical properties of the data of the Illustris simulation (Vogelsberger et al. 2014a, Genel et al. 2014), and the selection criteria that we have employed to ensure that the quantities that we examine resemble observables but without noise and systematics from observations. Then in section 3, we explain the methods for computing various one-point statistics of the spatial distribution of galaxies how we prepare our dark matter spatial data to resemble convergence maps. We show the statistical performance of the different summary statistics before we show the main results in section 4. In the discussion in section 5, we list the implications of our results and compare it to other simulations and observations. We also show how one may make use of the population offset statistical distribution from the Illustris data to construct a two-tail p-value test with a null hypothesis of $\sigma_{\rm SIDM} = 0$.

Our analysis makes use of the same flat Lambda Cold Dark Matter (ΛCDM) cosmology as the Illustris simulation. The relevant cosmological parameters are $\Omega_\Lambda = 0.7274, \Omega_m = 0.2726$, and $H_0 = 70.4$ km s^{-1} Mpc^{-1}.

2 THE ILLUSTRIS SIMULATION DATA

The Illustris simulation contains some of the most realistic, simulated galaxies to date, making it especially suitable for verifying the properties of galaxy clusters. We obtained our data from snapshot number 135 (cosmological $z = 0$) of the Illustris-1 simulation. The Illustris-1 simulation has the highest particle resolution and has incorporated the most comprehensive baryonic physics among the different Illustris simulation suites. The sophisticated galaxy formation model in Illustris-1 includes star formation rate, and stellar evolution due to environmental effects of the intracluster medium, such as ram pressure stripping and strangulation and feedback from Active Galactic Nuclei (AGN) etc. (Genel et al. 2014). The physics of stellar evolution were solved using a moving mesh code **AREPO** (Springel 2010). The observable properties of galaxies were statistically consistent with the Sloan Digital Sky Survey (SDSS) data (Vogelsberger et al. 2014a).

As the stellar population in Illustris were evolved from the initial condition, these makes the spatial distribution of galaxies in Illustris data more realistic than galaxies that are prescribed onto DM-only cosmological simulation data such as those used in Harvey et al. (2014). Gravitational effects in Illustris-1 have provided realistic dynamics and spatial distribution of subhalos. The simulated effects include tidal stripping, dynamical friction and merging. Since the profile of the galaxies clusters were not provided in symmetrical, parametric forms, we can study how asymmetry in the cluster profile affects the estimate of our summary statistic. This data allows us to examine cluster galaxies in a realistic, yet noise-free way. The softening length of the DM particles is 1.4 kpc and those of the stellar particles is 0.7 kpc, both in constant comoving units (Genel et al. 2014).

The two sets of data catalogs in use are obtained through two types of halo finders. The catalog that maps particles to the halo of a certain cluster was created by the **SUBFIND** algorithm. The friends-of-friends (FoF) finder (Davis et al. 1985) was further used to identify the affinity of galaxy-sized halos to a galaxy-cluster. These galaxy-size halos are referred to as *subhalos* and they are the dark matter hosts of what we refer to as galaxies in Illustris-1. Vogelsberger et al. (2014b) also extracted the absolute magnitude of each subhalo in the SDSS bands of g, r, i, z as part of the **SUBFIND** catalog using stellar population synthesis models.

For our analyses, we make use of galaxy clusters / groups with at least 50 member galaxies that are above a reasonable observation limit, i.e. apparent $i \leq 24.4$ which is the limiting magnitude of the DEIMOS spectrometer on the Keck telescope, when we assume a cosmological redshift of $z = 0.3$ in the i band. This is because of the relatively large statistical uncertainty if we try to analyze clusters with less than 50 member galaxies. As indicated by the right-hand panel of Fig. 1, a total of 43 clusters have survived this magnitude cut. These simulated galaxy clusters (or groups) have masses ranging from $10^{13} M_\odot$ to $10^{14} M_\odot$.

2.1 Cluster properties

2.1.1 *Relaxedness of the galaxy clusters*

Clusters undergo merger activities of a large range of physical scales and in the time scale of million of years. The dynamical history,

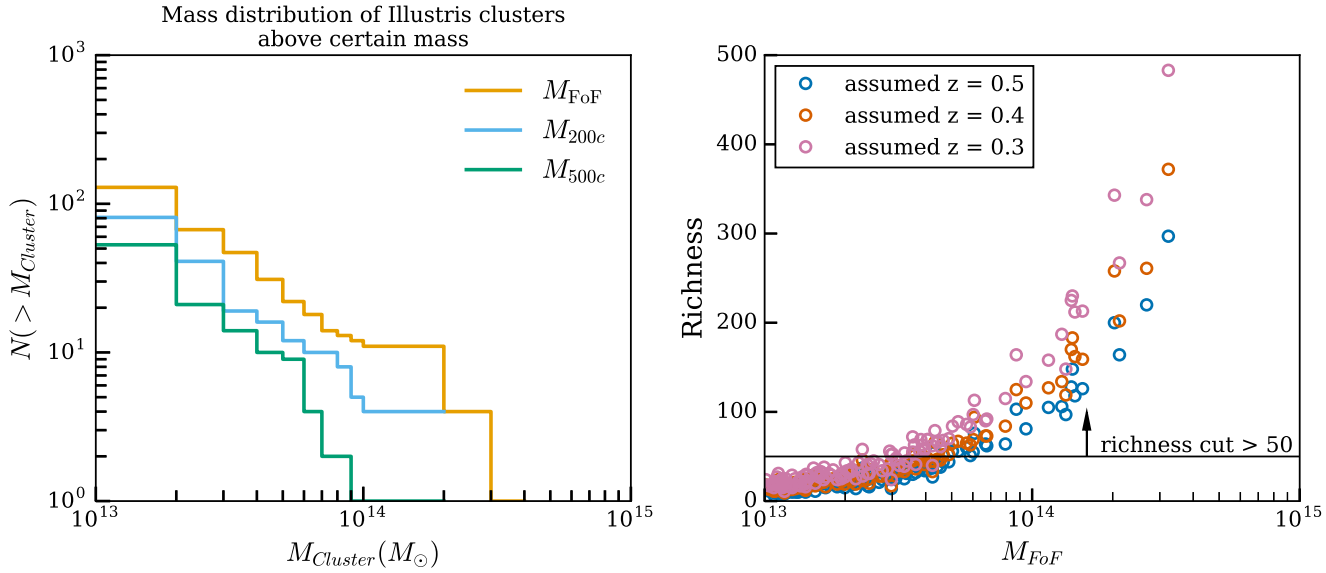

Figure 1. Left figure: Mass distribution of the group / cluster sized DM halos for different halo selection schemes. Mass estimates obtained by the FoF algorithm are labeled $M_{\rm FoF}$. Masses centered on the most bound particle within a radius those the average density is 200 or 500 times the critical density of the universe are labeled as M_{200c} and M_{500c} respectively. **Right figure:** Mass-richness relationship of galaxy clusters and groups with $M_{\rm FoF} > 10^{13} M_\odot$ assuming different cosmological redshifts of the observed clusters.

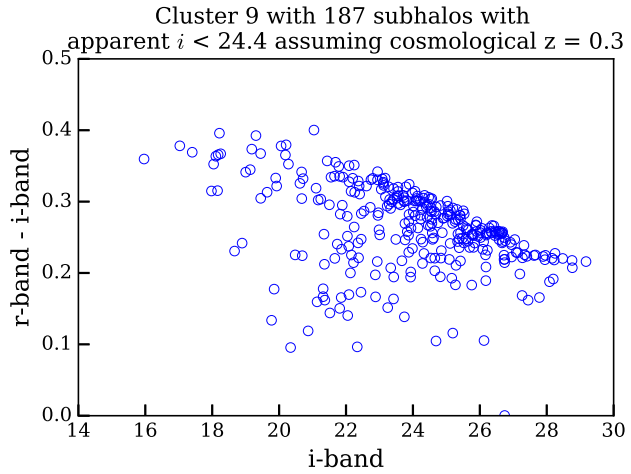

Figure 2. Color-magnitude diagram of one of the galaxy clusters that is selected for analysis. This cluster is the 9th most massive. The apparent magnitude is calculated assuming that the cosmological redshift (distance) is $z = 0.3$. We can see a clear overdense region that corresponds to a red-sequence. The color-magnitude diagrams of the other clusters can be found in the Jupyter notebook at https://goo.gl/TJmI6s.

or what we call "relaxedness" is hard to retrieve from simulations across different saved states and is missing from observations. We quantify the state of the cluster by providing several quantitative definitions of relaxedness and see how they correlate with Δs. Some definitions of relaxedness referred by the simulation community include:

- the ratio of mass outside the dominant dark matter halo over the total mass of the galaxy cluster
- the distance between the most bound particle from the center of mass as a function of R_{200c}.

which are computable from the Illustris data. To relate these simulation quantities, we compute more observation oriented quantities in the method section 3.0.2.

2.2 Selection of the field-of-view

We make use of the **SUBFIND** member particle for the DM and the **FoF** subhalo identification as our default volume selection scheme for each cluster / group. We understand that this choice of volume selection can be more ideal than observational conditions. We make use of this volume selection scheme for baseline comparisons.

Assuming a conservative line-of-sight (los) distance , i.e. cosmological redshift, with [TODO] $z = 0.4$, the projected extent for most of the Illustris galaxy clusters and groups, fits inside the field of view of telescopes, such as the Subaru Suprime Camera, which covers a physical area of [TODO] ~ 9 Mpc $\times 7$ Mpc. (See https://goo.gl/ClZNvM for a Jupyter notebook showing the extent of the Dark Matter distribution of the most massive 129 clusters)

2.2.1 *Spatial Projections*

The summary statistics are computed all based on 2D projections of the spatial location. For computing the summary statistic, we note that the order of projecting the data and taking the summary statistic is non-commutative. In order to represent the projection uncertainty, we compute even angular orientation as our line-of-sight by using HealPy, which is a Python wrapper for HEALPix[1] (Gorski et al. 2005). The number of projections that we employed is 384 for each cluster. Details of the implementation of the projection is available in Appendix A.

[1] HEALPix is currently hosted at http://healpix.sourceforge.net

Table 1. Selection criteria for stellar subhalos (member galaxies) for each cluster / group

Data	Selection strategy	Sensitivity	Relevant section
Field of view (FOV)	FoF halo finder	comparable to FOV of the Subaru Suprime camera	2.2
Observed filter	i-band	consistent among the redder r, i, z bands	2.3
Cluster richness	$i \leq 24.4$ and $z = 0.3$	sensitive to the assumed cosmological redshift of cluster and the assumed limiting magnitude of telescope	2
Two-dimensional projections	even HEALPix samples over half a sphere	discussed as results	2.2.1

2.3 Properties of the galaxies in Illustris clusters

Different galaxies have different masses, so they should not be considered with equal importance for peak identification, which requires summing the mass proxies of different galaxies. One of the most common weighting schemes employed for galaxy data is to weight by the luminosity in a particular band. We make use of the $i-$band magnitude associated with each subhalo as the weight. Since the $i-$band is one of the redder bands, the mass-to-light ratio is not skewed as much due to star formation activities. We further examined if the colors distribution of galaxies in Illustris-1 are similar to the observed color-magnitude diagrams for clusters. The Illustris cluster galaxies are realistic enough that it is easy to identify an overdense region of galaxies known as the red-sequence in the color-magnitude diagram such as Fig. 2. The red-sequence is prominent even if we use other colors formed by different combinations of r, i, z bands.

3 METHODS

A common and the most precise way of summarizing the DM distribution in a galaxy cluster is by finding the lensing peaks (Medezinski et al. 2013, Markevitch et al. 2004, Zitrin et al. 2013). Additionally, the peak region is physically interesting due to the higher particle density and interaction rates. The most direct analogous statistic for summarizing the member galaxy population in a cluster is therefore, also the peak. Comparing the DM peak with the summary statistics of the galaxy population that are not the peak therefore can have an *offset* purely due to the difference in the choice of the statistic for summarizing the two sets of identically distributed data.

We compare four common point statistic or location for summarizing the member galaxy population in a galaxy cluster:

(i) Weighted centroids
(ii) Weighted density peak via density estimate
(iii) Shrinking aperture estimate
(iv) Brightest cluster galaxy (BCG)

We avoid any manual methods for comparison purposes, scalability and reproducibility. Since all the methods listed in this paper are automated with the source code openly available, it is possible for future studies to reuse our code for comparisons. Furthermore, a major advantage for automation is that it allows us to apply the same methods across the different snapshots of the (Illustris) simulations to examine the variability of Δs across time.

3.0.1 *Computing the weighted centroid*

We follow the usual definition of the weighted centroid is just:

$$\bar{\mathbf{x}}_w = \frac{\sum_i w_i \boldsymbol{x}_i}{\sum_i w_i}, \tag{3}$$

with $\boldsymbol{x}_i$ being the positional vector of each subhalo and we use the i-band luminosity as the weight w_i for the i-th galaxy. Centroids can be biased 1) by subcomponents from merging activities yet the centroid estimate do not provide explicit evidence for ongoing merger or accretion. These estimates are also sensitive to odd boundaries of the field of view.

3.0.2 *Cross-validated Kernel Density Estimation (KDE) and the peak finder*

Finding the exact peak of a sets of data points involves computing the density estimate of the data points and sorting through the density estimates. A specific version of this density estimation process is known as histogramming. During the making of histogram, each data point is given some weight using a tophat kernel and the weights are summed up at specific data locations (e.g. $\mathbf{x_i}$). Histogram is not good for peak estimate for *sparse* data for two reasons: 1) the choice of laying down the bin boundaries strongly affects the count in each bin, 2) the choice of bin width also strongly affects the count. Only when the available number of data points for binning is large, the estimates of histograms and smoothed density estimates are approximately the same. The number of member galaxies (¡ 500) is sparse enough for the uncertainty introduced by histogramming to bias our peak estimate. For the density estimate of galaxy luminosity, we adopt a Gaussian kernel. The exact choice of the functional form of the smoothing kernel does not dominate the density estimate as long as the chosen kernel is smooth (Feigelson & Babu 2014).

The most important parameter of computing the density estimate is the bandwidth of the smoothing kernel, which takes the form of a matrix in the 2D case. We illustrate the choice of kernel width with Fig. 3. When the kernel width is too large (bottom left panel), the data is over-smoothed, resulting in a bias of the peak estimate. On the other hand, when the kernel width is too small, it results in high variances of the estimate and result in too many peaks due to noise. The decision of having to balance between creating high bias or high variance estimates is also known as the bias-variance tradeoff.

A well-known way to minimize the fitting error from the density estimate is through a data-based approach called cross-validation to obtain the optimal 2D smoothing bandwidth matrix (H) of the 2D Gaussian kernel for the density estimate $\hat{f}$:

$$\hat{f}(\chi; H) = \frac{1}{n} \frac{1}{(2\pi)^{d/2}|H|^{1/2}} \sum_{i=1}^{n} w_i \exp((\chi - \mathbf{x}_i)^T H^{-1} (\chi - \mathbf{x}_i)), \tag{4}$$

where the dimensionality is $d = 2$ for our projected quantities, χ represents the uniform grid points for evaluation, and $\mathbf{x_i}$ contains the spatial coordinates for each of the identified member galaxies that survived our brightness cut and w_i is again the $i-$band luminosity weights for each galaxy. The idea behind cross-validation is to leave a small fraction of data point out as the test set, and use the rest of the data points as the training set for computing the estimated density.

Figure 3. This figure is adapted from VanderPlas et al. 2012 from http://www.astroml.org/book_figures/chapter6/fig_hist_to_kernel.html under the fair use of the BSD license.

Then it is possible to minimize the asymptotic mean-integrated squared error (AMISE) by searching for the best set of bandwidth matrix values, eliminating any free parameters.

Specifically, we made use of the smoothed-cross validation (Hall et al. 1992) bandwidth selector in the statistical package **ks** (Duong 2007) in the **R** statistical computing environment (R Core Team 2014). Among all the different **R** packages, **ks** is the only package capable of handling the magnitude weights of the data points while inferring the density estimates (Deng & Wickham 2011). Although the particular implementation of KDE has a computational runtime of $O(n^2)$, the number of cluster galaxies is small enough for this method to finish quickly.

After obtaining the KDE estimate, we employed both a first and second-order finite differencing algorithm to find the local maxima. The local maxima were then sorted according to the KDE density in a descending fashion before we perform peak matching and compute the offset. The exact procedure is discussed in section 3.3.

3.0.3 Shrinking aperture estimates

Another popular method among astronomers for finding the peak of a spatial distribution is what we call the shrinking aperture method. While we do not endorse this method, we test if the shrinking aperture method is able to reliably recover the densest peak. This method is dependent on the initial diameter and the initial center location of the aperture. This method does not evaluate if the cluster is made up of several components. The estimate using the shrinking aperture algorithm can be biased by substructures. The only way to inform the algorithm about substructures would be to introduce another parameter to restrict the center of the aperture, or to partition the data with another (statistical) algorithm. Furthermore, the convergence rate for this iterative algorithm is not analytical and is dependent on both the data and the parameters. We use a convergence criteria of having the aperture distance not change more than 2% between successive iterations as a reference. The actual implementation in Python can be found at https://goo.gl/nqxJl8 while the pseudo-code can be find in Appendix B.

3.0.4 Brightest Cluster Galaxies (BCG)

The BCGs are formed by the merger of many smaller galaxies. The galaxy-cannibalism makes BCGs typically brighter than the rest of the cluster galaxy population by several orders of magnitude. However, star formation can cause less massive galaxies to be brighter in the bluer photometric bands. To avoid star formation from biasing our algorithm for identifying the BCG, we find the brightest galaxies in redder bands i.e. the r, i, z bands and found that they give consistent results for all selected clusters. We used the $i-$band to pick the BCG for the plots and the final results.

3.1 Comparison of the methods from Gaussian mixture data

In order to examine the statistical properties of commonly used point-estimates of the distribution of the galaxy data, we test them on data drawn from Gaussian mixtures with known mean and variance. (See Fig. 4). The main factors that affect the performance of the methods are sensitive to the statistical fluctuations of the drawn data, e.g. the spatial distribution of the data, including 1) the density profile and 2) the location(s) of subdominant mixtures, , and 3) the number of data points that we draw. It is also not enough to just compare the performance by applying each method for one realization of the data. We provide the 68% and the 95% confidence regions by applying the each method for many Monte Carlo realizations. In general, the peaks identified from the KDE density is closer to the peak of the dominant mixture (more accurate) than both the weighted centroid method and the shrinking aperture method. For example, in the bottom middle panel, it is clear that the green contours that represents the confidence region for the shrinking aperture peak is biased due to the substructure, whereas the confidence region for the centroid is so biased that it is outside the field of view of that panel. For the bottom right plot, there is also a catastrophic outlier for the shrinking aperture method for 500 data points. The outlier shows how the shrinking aperture method can have radical behavior when there are subclusters in the data.

3.2 Modeling the DM map in Illustris-1 and the lensing kernel

The most well established method of inferring the projected dark matter spatial distribution from observations is through gravitational lensing. It works by detecting subtle image distortions of background galaxies due to the foreground dark matter. The resolution of the inferred map therefore depends on the properties of the source galaxies that are being lensed, such as the projected number density, intrinsic ellipticities and morphology etc. To achieve a sufficient signal-to-noise ratio for lensing, Hoag et al. 2016 has performed simulation for inferring the optimal size for a Gaussian smoothing kernel for the cluster MACSJ0416. In the strong lensing regime, Hoag et al. found a resolution of 11 arcseconds can best fit the data. This kernel translates to a physical size of 50 kpc assuming a cosmological redshift of $z \approx 0.3$. To compute a DM spatial distribution, we first make histogram with 2 kpc $\times$ 2 kpc bin size which is larger than the DM softening length of 1.4 kpc. After that, we use a Gaussian smoothing kernel of the DM histogram Illustris DM particle data. We do not perform a cross-validated KDE that has $O(n^2)$ runtime on the DM data because the number of DM particles for each cluster is of [TODO double check] the order of millions. The DM resolution is high enough for the histograms to be accurate. Physically, the histograms of the dark matter of each cluster is analogous to a convergence map from a lensing analysis.

3.3 Finding the offsets and its population distributions

It is possible to have several peak estimates from the KDE of the galaxy population. From the density estimate at each peak, we can sort the peaks according to their densities. We only match the peaks that are at least 20% as dense as the densest galaxy-luminosity peak to avoid computing the offsets of negligible substructures, such as the peaks due to galaxies that are located far away from the main concentraiton of mass.

In general, there are many more DM peaks because there are many more dark subhalos than galaxies for each cluster. To find the nearest match to the significant galaxy peaks, we construct a k-dimensional tree (KD-Tree) using the densest $n_{\rm DM}$ number of

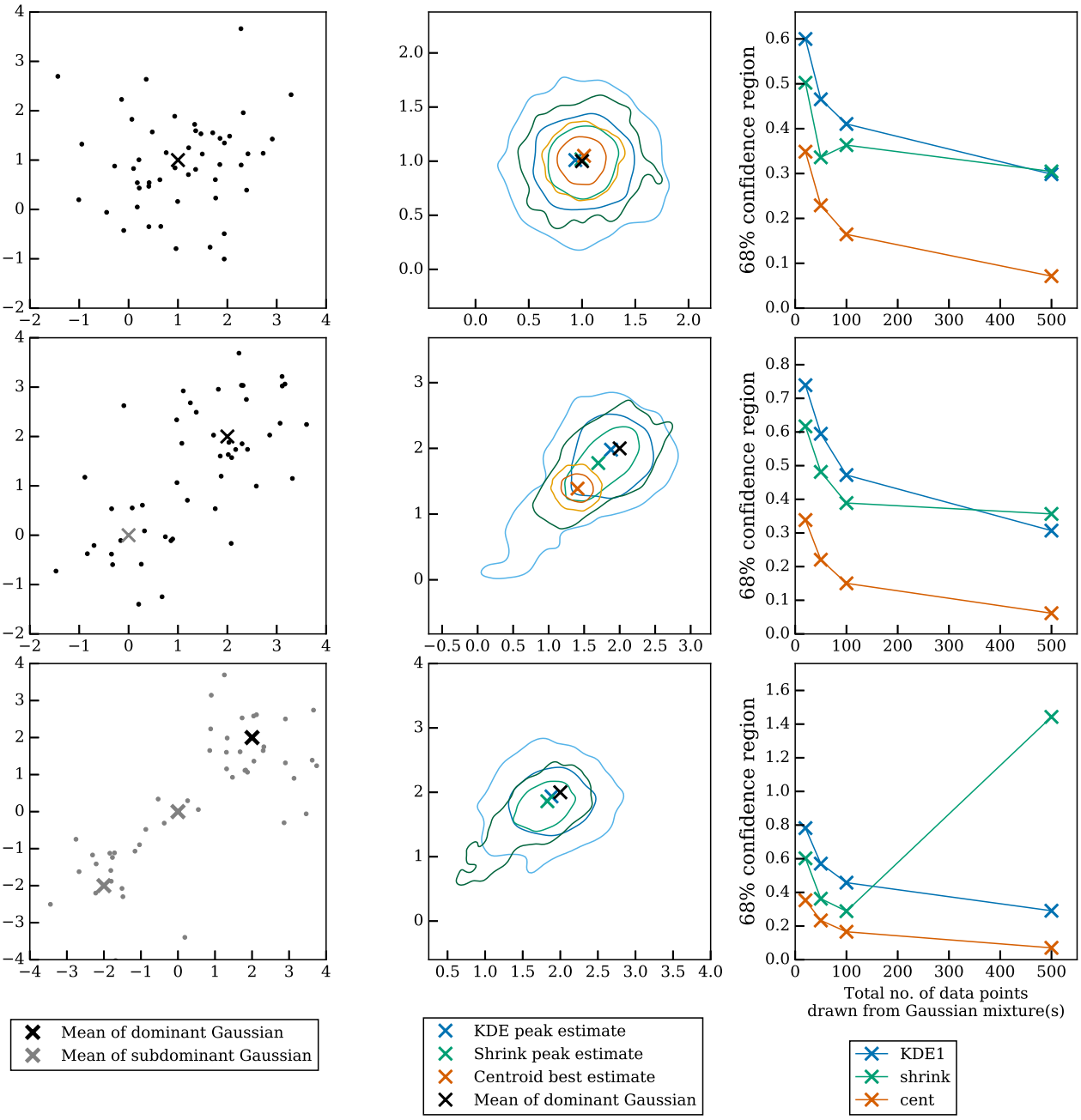

Figure 4. Comparison of peak finding performances of different methods by drawing data points (i.e. 20, 50, 100, 500) from known number of Gaussian mixtures. Panels from the top row contain data drawn from a single Gaussian mixture. The panels from the middle row contain data from two Gaussian mixtures with weight ratio = 7:3. The panels from the bottom row contain data drawn from three Gaussian mixtures with weight ratio = 55:35:10. The left column shows how 50 data points drawn from the fixed number of Gaussian mixtures look like. Due to the statistical nature of this exercise, we sampled the data and performed the analyses [TODO: state how many times] many times to create the (68% and 95%) confidence contours of the estimates in the zoomed-in view of the data in the middle column. The rightmost column shows how the size (median contour radius) of the confidence regions vary as a function of the number of drawn data points from the Gaussian mixtures. From the middle and the rightmost column, we can tell that the KDE peak estimate is the most accurate but less precise for estimating the sampled data from each set of data.

DM peaks:

$$n_{\rm DM} = \begin{cases} 3 \times (n_{gal} + 1) & \text{if } n_{gal} < 3 \\ 3 \times n_{gal} & \text{if } n_{gal} \geq 3. \end{cases} \quad (5)$$

where n_{gal} is the number of significant galaxy peak, and $n_{\rm DM}$ is the number of peaks that went into the construction of the KD-tree. During the matching, we have $n_{\rm DM} > n_{\rm gal}$. When there are more than one dense galaxy peaks located far away from one another, the top few densest DM peaks can be located around the same galaxy peak. i.e. there is no one-to-one matching between the luminosity of galaxies and the density of detected DM peaks. Matching purely based on density and luminosity leads to larger offsets. From inspection, using eq. (5) works well to match the appropriate peaks.

After matching the peaks, we aggregate the distribution of

offsets for all the clusters. The width of the marginal distribution $P(|\Delta s|)$ represents the uncertainty both from the different cluster masses, dynamical states and the projections. We provide the best estimated location of $P(|\Delta s|)$ using the biweight statistic (Beers et al. 1990). Note that for

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \tag{6}$$

where Δx and Δy are the two projected spatial offsets by subtracting the DM peak coordinates from the galaxy peak coordinates. We discuss the possible loss of information and the caveats of the variable transformation in eq. (6). The biweight estimate will not center on zero even if Δ

[TODO] make sure that the plots reflect this.

[TODO] talk about how I computed the offset / offset distribution for the KDE peaks

The inference of the estimated value and the confidence intervals will be incorrect after taking the absolute value of the offsets. The projected offsets for different clusters and projections give us as estimate of the population distribution of the offsets.

[TODO] talk about how I computed the other offsets, BCG / centroid / shrinking aperture etc.

4 RESULTS

We summarize the main results here and leave the detailed tables of results in Appendix E.

4.1 Galaxy-DM Offset in Illustris

4.1.1 Two-dimensional(2D) offsets

From Fig. 6, we can tell that the population estimates.

The variance of the offsets due to projection is [TODO]. In particular for the most massive galaxy clusters with $10^{14} M_{\odot}$

- those between BCG, the most bound particle and the other masses.
- explain the variation of the offsets for the same cluster under different projections

While there is a tight distribution for $\Delta s_{\rm BCG}$ that peaks around zero, there is a non-negligible population of data points that lies outside this peak. [TODO] investigate what is WRONG with those outliers.

4.2 Correlations between different variables and the offsets

We subset the data and visually inspected the samples with the largest galaxy-DM offsets $\Delta s_{\rm KDE}$. We found that ...

4.2.1 Correlations between the offsets and properties of the cluster / groups

[TODO] examine the relationship between

- 3D relaxedness
- mass
- richness

5 DISCUSSION

[TODO]Density estimate based only on galaxy number density without luminosity weights do not resemble the DM distribution.

It is not easy to compare the result of this study to other study due to the differences in the multi-step method for inferring the "peak", or the "centroid".

5.1 Other findings from the visual inspection of the simulated galaxy clusters

From the high resolution visualization of the DM maps (two-dimensional histograms with 2 kpc bins) in Fig. 5, we can tell that some clusters clearly possess multiple subclusters with visible separations between the subclusters, e.g. cluster 12 visualized on the panels on row 3 of the plot. However, there are also many clusters that contain only one main component but show several closely-located density peaks in the densest region. This illustrates why finding a 'center' or peak of a cluster is an ill-defined notion. Most of the density-based statistics are only unambiguous for smooth, symmetrical data.

Furthermore, we show a lower-resolution visualization of the DM surface map in Fig. [TODO] compared to the higher resolution map, we can see a clear shift in the peak location. This illustrates that peak / center finding is also subject to noise from the data.

Sometimes when the peaks are mismatched, it is due to a resolution difference between the higher resolution DM data and the sparse galaxy data. (See middle bottom panel in Fig. 5)

The main reasons for the existence of several peak estimates include: 1) Extended substructures that exist within the cluster, which we call subclusters, although the boundaries of subclusters are ill-defined. 2) Galaxies that are located further away from the main concentration of mass. These peaks are excluded based on density cuts. 3) This last case illustrates the difficulty of pinpointing a point-estimate such as a 'center' . It is possible to have more than one bright galaxies spatially close to the main concentration of mass. (See the right most panel of Fig. 5 where the galaxy luminosity peaks are represented with squared markers that are colored according to the relative density to the densest galaxy peak of that projection).

5.2 The statistical performance of different peak finding methods

The difficulty in matching is not necessarily due to wrong algorithms. When the substructures of several important physical scales are present due to hierarchical clustering, peak finding is not a well defined problem. It is possible for a smaller group-sized halo to be embedded in a galaxy-cluster size halo, or several group-sized halo have the majority of its mass overlapped with another halo. There is no unique solution for specifying the physical boundaries of substructures before finding the corresponding peaks. While the peak-matching by galaxy-peaks that we performed is an ad-hoc approach, only picking spherically-looking can only guarantee relaxedness up to a certain extent. The number of non-relaxed clusters far outnumber the ones that are dynamically relaxed.

The situation is subsectionComparison to other simulations

5.2.1 Comparison to other cosmological simulations

Assume dynamical equillibrium

Cui et al. (2016) X-ray center, BCG

does not provide the pairwise offset between the DM and the

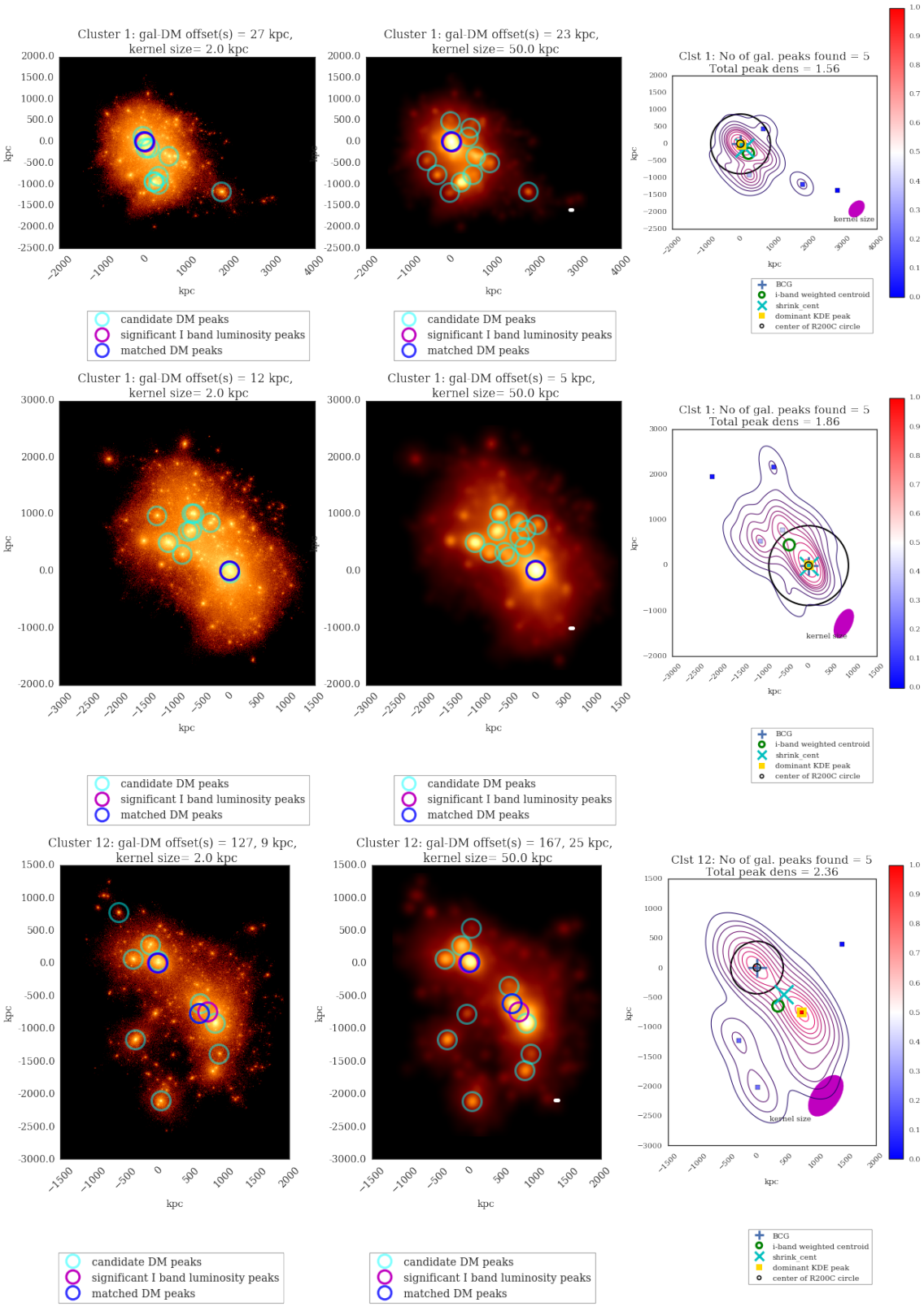

Figure 5. [TODO merge margin between left and middle panels] Visualization of clusters (each row is for the same projection of the same cluster). **Left column:** Projected density distribution of DM particle data (density overlay). The identified density peaks are indicated by colored circles. **Middle column:** The same DM projection but with treated with a 50 kpc smoothing kernel (kernel size indicated by white dot on lower right of the figure. Note that the thickness of the dot may be larger than 2 kpc for the plots on left hand column). **Right column:** Projected galaxy kernel density estimates (KDE) of the i-band luminosity map for the member galaxies of the same clusters. Each colored contour denotes a 10% drop in density mass starting from the highest level in red. Each of the magenta ellipse on the bottom right corner of each plot show the Gaussian kernel matrix H from eq. (4). The big black circle is centered on the most bound particle as identified by **`SUBFIND`** and the radius of the circle indicates the three-dimensional region in which the average density is 200 times the critical density of the universe (a.k.a. R_{200C}). See http://goo.gl/WiDijQ and http://goo.gl/89edcM for the visualization of the selected clusters inside two Jupyter notebooks.

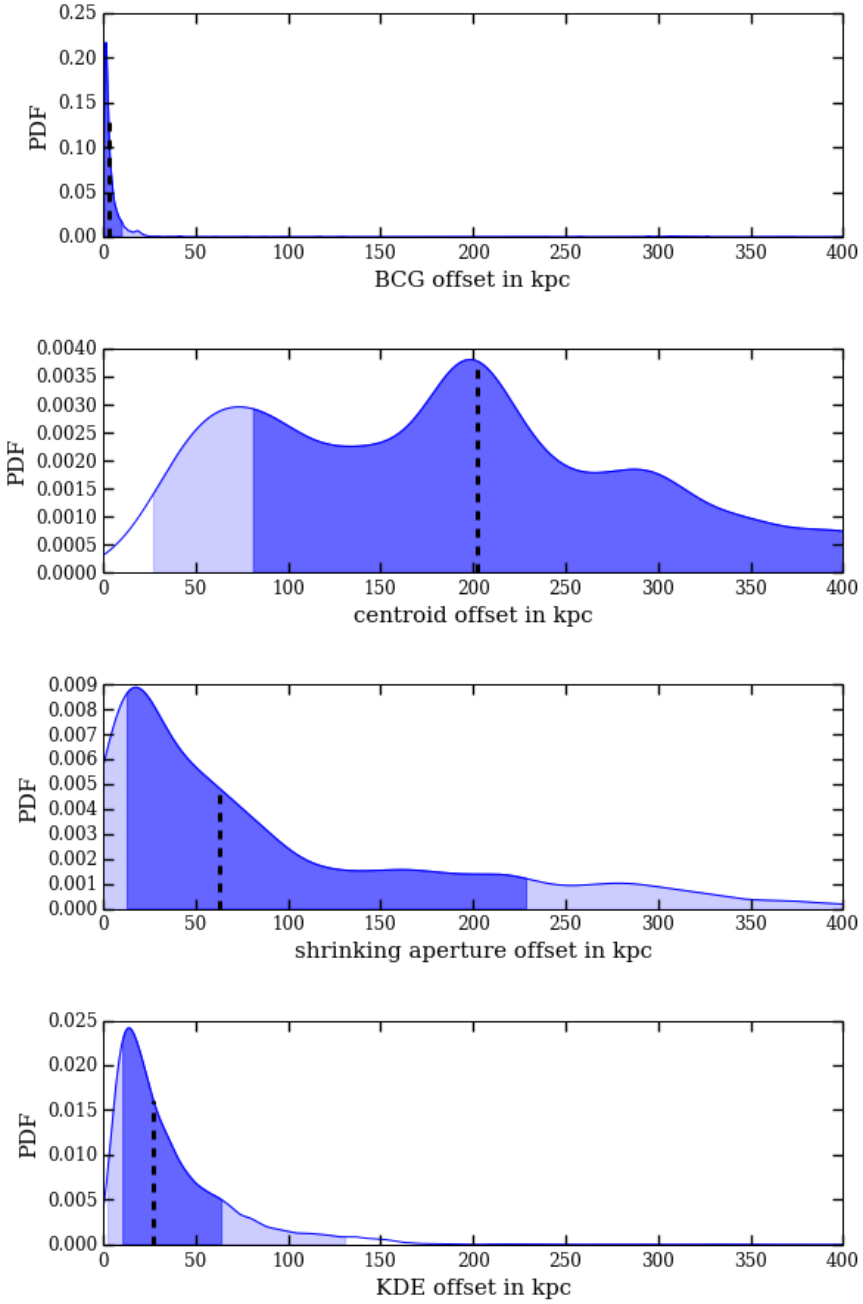

Figure 6. The distribution of different offsets of [TODO] clusters with [TODO] projections. The dark blue area indicates the 68% confidence interval while the light blue area shows the 95% confidence interval. We plot the offset after taking the absolute magnitude, which The estimates from the absolute magnitude of the offsets are pushed towards larger values.

galaxy distributions but rather the offset between the minimum potential and DM. computing the offset from the population statistic of the DM spatial distribution with the population galaxy distribution destroys the possible correlations.

Fig. 3 in Cui et al. (2016) shows both the DM peak and the galaxy offsets to the minimum gravity potential center but that representation destroys the correlations.

kernel smoothing approach use minimum potential position as a proxy for the maximum SPH density

Voronoi Tessellation Density (VTD)

gives no confidence estimate for comparison. width of histograms can affect the distribution taking the absolute magnitude of the offset can bias the data Harvey et al. (2014) consistent with zero offsets using the **`Wavedetect`** wavelet analysis algorithm. Galaxies prescribed to the DM simulation and not simulated to see how the surrounding environment

Provided nice theoretical motivation for describing how galaxies and DM population would behave They have only verified that their model of SIDM work for $\sigma_{\rm SIDM} = 0$ but not any model with $\sigma_{\rm SIDM} > 0$

5.2.2 *Comparison to other staged simulations*

In this work, we test purely for the offset due to statistical noise and missing variables such as projection and impact parameter

Randall et al. (2008) can be thought of a controlled experiment where only the effects of SIDM is in place. This is because a large number of tracer galaxies were painted on as collisionless particles, using identical distributions between the galaxies and the DM particles at the beginning of each run.

Statistical uncertainties overwhelm the signal of SIDM.

Robertson et al. (2016)

[TODO] cite Stacey

The

It is not clear about the ability of $\Delta s_{\rm SIDM}$ for differentiating between different $\sigma_{\rm SIDM}$ models.

It is unclear that Δs is a sensitive observable to signatures of $SIDM$.

At $\sigma_{\rm SIDM} \approx 1.25 \mathrm{cm}^2$ / g, Randall et al. 2008, Robertson et al. 2016 only found an offset of $50 kpc$, this is completely within the population uncertainties of offsets, while the projection uncertainty of a cluster can be as large as [TODO] kpc.

Bullet cluster

5.2.3 *Estimation of the SIDM cross section from the Illustris offsets*

Since the Illustris simulation assumes CDM, it is interesting to see if we can infer any non-zero $\sigma_{\rm SIDM}$ from the offset estimates based on different methods.

5.3 Comparison to other observational studies

Weak lensing observations infer mass distribution of both baryonic and dark matter.

Central galaxy paradigm (CGP)

- Ford et al. (2014) paper about miscentering in CFHT
- George et al. (2012) about miscentering

There are many aspects of the analysis that is not covered by this study that are performed for analyzing observational data, such as

- galaxy membership identification along the line of sight
- removal of foreground galaxies

that are important for calculating the $\sigma_{\rm SIDM}$ with using a galaxy-DM offset.

5.4 Galaxy-DM Offset in observations of Merging Galaxy Clusters

how the offsets will translate to a $\sigma_{\rm SIDM}$.

5.5 How to use Δs to constrain $\sigma_{\rm SIDM}$

As a faster, preliminary step to show that σ_{SIDM} is plausible, one can compute the distribution of offsets in the Illustris simulation as a null hypothesis. We therefore provide the confidence region estimates in E1.

Statistical aspects that one should take into account when constructing such a test include:

- whether the population PDF of offsets have converged

When there are real observations that lie sufficiently far away from the 95% confidence region of the population distribution of the Illustris offsets, it is less likely that the null hypothesis that $\sigma_{\rm SIDM}$ is true.

However the p-value can only tell the following: given there is $\sigma_{\rm SIDM} = 0$, how (un)likely it is to see the level of Δs that we see.

To properly infer $\sigma_{\rm SIDM}$ as a parameter estimation problem and account for various uncertainties, one should produce a simulation suite with a fixed $\sigma_{\rm SIDM}$ and compute the marginal likelihood of fitting the offset distributions with observations. Proper marginalization of nuisance parameters may include projections, mass and relaxedness etc.

Alternative observables such as Bulleticity as suggested by Massey et al. (2011) should be verified with simulations those $\sigma_{\rm SIDM} > 0$.

Extremely computationally intensive because nobody has shown that painted on galaxy population is realistic enough that the physical effects are sufficient.

Need to show the validity of ignoring projection in cosmological simulations with non zero $\sigma_{\rm SIDM}$.

Machine learning methods to paint galaxies to DM halos http://arxiv.org/abs/1510.07659

6 SUMMARY

We showed that

- the peak finding method from KDE for the density of cluster galaxies is the second least biased due to substructures from our test data.
- the observed offsets from various merging galaxy clusters are not statistically significant when compared to the offsets from a ΛCDM simulation
- all existing peak finding methods have non-negligible uncertainty due to the small number of data points. When dealing with small number of cluster samples, the uncertainties of the peak locations should not be ignored.
- the resolution of the DM distribution can affect individual estimates Δs but do not show significant bias for the population estimate. However, the lower the resolution, the higher the variance of the population estimate.

Furthermore, we have provided a set of python functions for making accurate contour levels for spatial maps and inferring density peaks at https://goo.gl/MNrSQV.

While this paper does not provide a solution to the complete statistical model for galaxy clusters, it points out some of the aspects that good models should incorporate, especially when the utility of studying the cluster is to constrain $\sigma_{\rm SIDM}$.

Aspects that should be handle with care include:

- the mass profile, especially when the cluster is multimodal
- the evaluation of galaxy membership and the corresponding luminosity peak(s)

7 ACKNOWLEDGEMENTS

KN would like to thank Professor Thomas Lee for the helpful discussion of the construction of the Monte Carlo p-value hypothesis test. Part of the work before the conception of this paper was discussed during the AstroHack week 2014. KN would like to thank

Phil Marshall and Jake Vanderplas for preliminary discussions for analyzing galaxy clusters.

Part of this work was performed under HST grant (TODO ask Dave for grant number).

REFERENCES

Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Bradač M., et al., 2006, ApJ, 652, 937
Cui W., et al., 2016, MNRAS, 456, 2566
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dawson W. A., 2013, PhD thesis, University of California, Davis
Deng H., Wickham H., 2011, Technical report, Density estimation in R. had.co.nz
Duong T., 2007, J. Stat. Softw., 21, 1
Feigelson E. D., Babu G. J., 2014, Contemp. Phys., 55, 126
Ford J., et al., 2014, MNRAS, 447, 1304
Genel S., et al., 2014, MNRAS, 445, 175
George M. R., et al., 2012, ApJ, 757, 2
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hall P., Marron J. S., Park B. U., 1992, Probab. Theory Relat. Fields, 92, 1
Harvey D., et al., 2014, MNRAS, 441, 404
Hoag A., et al., 2016, arXiv Prepr., p. 48
Jee M. J., et al., 2015, ApJ, 802, 46
Kahlhoefer F., Schmidt-Hoberg K., Frandsen M. T., Sarkar S., 2013, MNRAS, 437, 2865
Markevitch M., Gonzalez a. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819
Massey R., Kitching T., Nagai D., 2011, MNRAS, 413, 1709
Medezinski E., et al., 2013, ApJ, 777, 43
R Core Team 2014, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/
Randall S. W., Markevitch M., Clowe D., Gonzalez A. H., Bradač M., Bradac M., 2008, ApJ, 679, 1173
Robertson A., Massey R., Eke V., 2016, arXiv Prepr., p. 20
Springel V., 2010, MNRAS, 401, 791
VanderPlas J., Connolly A. J., Ivezic Z., Gray A., 2012, in 2012 Conf. Intell. Data Underst.. IEEE, pp 47–54, doi:10.1109/CIDU.2012.6382200, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6382200
Vogelsberger M., et al., 2014a, MNRAS, 444, 1518
Vogelsberger M., et al., 2014b, Nature, 509, 177
Zitrin A., Menanteau F., Hughes J. P., Coe D., Barrientos L. F., Infante L., Mandelbaum R., 2013, ApJ, 770, L15

APPENDIX A: GETTING UNIQUE 2D PROJECTIONS OF THE CLUSTERS

In 3-dimensional space, rotation operations are non-commutative. We first actively rotate our clusters by the azimuthal angle ϕ before we rotate the particle according to the elevation angle ξ. Then we project to the transformed x-y plane. With this rotation scheme, two projections are identical if

$$|\phi_1 - \phi_2| = \pi \tag{A1}$$

It represents viewing the same cluster from opposite sides. To save disk space, we only compute the statistics from one of the two unique projections.

APPENDIX B: ALGORITHM OF THE SHRINKING APERTURE ESTIMATES

Data: subhalo that satisfy cuts as a galaxy

initial aperture centroid = mean galaxy location in each spatial dimension
distance array = euclidean distances between initial aperture center and each galaxy location
aperture radius = 90th percentile of the distance array
while *(newCenterDist - oldCenterDist) / oldCenterDist* $\geq$ *2e-2* **do**
  new data array = old data array within aperture
  newCenter = mean value of new data along each spatial dimension
end

Algorithm 1: Shrinking aperture algorithm

APPENDIX C: TRANSFORMATION OF VARIABLE THAT REPRESENTS THE OFFSET

Taking the absolute value of the offset represents a variable transformation of

$$s \rightarrow |s|, \tag{C1}$$

which does not have a one-to-one mapping so we cannot perform transformation of variables via the computation of the Jacobian term that one normally would perform for

$$x \rightarrow y \tag{C2}$$

$$p(y)dy = p(y(x))\frac{dy}{dx}dx \tag{C3}$$

where $p(y)$ is the probability density function (PDF). Instead, to compute the PDF correctly under the operation of taking absolute magnitude, we make use of the fact that the cumulative distributions:

$$P(|S| \leq s) = P(-s \leq S \leq s) \tag{C4}$$

APPENDIX D: CORRECT COORDINATE TRANSFORMATION FOR THE 2D OFFSETS

Care must be taken when trying to With symmetry arguments, we know that we can represent the offsets in polar coordinates:

$$\Delta \boldsymbol{s} = (\Delta s, \theta) \tag{D1}$$

[TODO] These 2D offset estimates can take both positive and negative values because of the different spatial configurations of the galaxies and the dark matter. Naively taking an absolute value of the population of offset to compile the distribution , i.e.

$$s \rightarrow |s| \tag{D2}$$

is inappropriate because the mapping in eq. (D2) is not one-to-one. If the variable transformation is done correctly as outlined in [TODO] Appendix,

Table E1. Offsets for the full sample of 43 clusters at 2 kpc resolution.

Offset (kpc)	Location	68% CI†	95% CI†
BCG	-0.0	-7.486 , 10.35	-123.4 , 848.7
Weighted centroid	-37.19	-344.8 , 246.1	-806.0 , 567.9
Shrinking aperture	-7.191	-134.0 , 113.7	-368.4 , 343.8
KDE	-0.0	-38.38 , 33.79	-110.5 , 82.13

† CI stands for the confidence interval centered on the biweight location estimate, with 68% of the probability density contained in the 68% confidence interval.

Table E2. Absolute offsets for the full sample of 43 clusters at 2 kpc resolution.

Offset (kpc)	Location	68% CI†	95% CI†
BCG	9.353	6.337 , 19.82	-2.656 , 1292.0
Weighted centroid	195.2	86.41 , 402.7	10.59 , 797.5
Shrinking aperture	70.07	13.44 , 236.4	-29.63 , 446.7
KDE	21.7	6.886 , 59.93	-3.722 , 121.3

† CI stands for the confidence interval centered on the biweight location estimate, with 68% of the probability density contained in the 68% confidence interval.

APPENDIX E: TABLE OF RESULTS

.

This paper has been typeset from a TEX/ LATEX file prepared by the author.