Mon. Not. R. Astron. Soc. **000**, 1–7 (2015) Printed 28 December 2015 (MN LaTeX style file v2.2)

Galaxy-dark matter offsets in galaxy clusters and groups of the Illustris simulation

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

arXiv

ABSTRACT

Galaxy clusters, which mainly compose of dark matter (DM), can be rare test beds for the particle properties of DM. However, the continuous merger and accretion events of clusters also complicate the modeling of galaxy clusters. With the various modeling choices and observational constraints, we need to carefully account for the uncertainties for us to give meaningful quantitative constraints from the studies of galaxy clusters. In this paper, we test various summary statistics of the different components of galaxy clusters by applying them to data from a cosmological simulation, the Illustris simulation.

quantify the uncertainties of the different one-point statistics that are intermediate quantities for characterizing the spatial offsets between the member galaxies and the dark matter components.

Key words: galaxy clusters, dark matter, something else

1 INTRODUCTION

Galaxies that belongs to a galaxy cluster or group are stochastic samples of the underlying dark matter (DM) mass density. We quantify the bias and uncertainty associated with the one-point summary statistic for summarizing the physical state of a galaxy cluster.

Uncertainties affect the conclusion for the computation the hypothesis test / parameter estimation Previous work on quantifying galaxy-DM offsets included What centroids they have used 1) papers reported using centroids but did not state what centroid that they used 2) papers used center of mass 3) papers that used peaks

In this paper, we 1) extract realistic observables from the Illustris simulation for comparison with observations, 2) identify practical, objective one-point 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. We call this offset as

$$\Delta s \equiv |\mathbf{s_{gal}} - \mathbf{s_{DM}}|, \tag{1}$$

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. Finally, we provide the distribution and investigrate origin of Δs.

The data that we use: The Illustris simulation one of the most detailed cosmological simulations of our universe. Cold Dark matter Our paper adopt the same cosmology

The highest resolution Illustris-1 simulation. With a snapshot at $z = 0$. Two sets of halo finder - one is for halo sized stuff the other one for particles Subfind

Terminology halos - what we refer to cluster sized halos identified by RockStar / Subfind

subhalos - galaxy sized observation bands - u, g, r, i, z

2 THE ILLUSTRIS SIMULATION DATA

2.1 Galaxy Clusters and the Member Galaxy Sample

The evolution of the baryonic component of the Illustris simulation was simulated in a self-consistent manner by solving the Euler equation with the moving-mesh technique AREPO **?**. More details about the galaxies can be found in **?**. This results in galaxies with well-resolved morphologies, such as early-type, late-type and irregular galaxies. A galaxy formation recipe in Illustris implies that, most, if not all subhalos contain baryonic stellar components, but not all the subhalos contain enough stellar content to be bright enough as galaxies.

2.2 Member galaxy selection criteria

We only make use of subhalos in a cluster that are brighter than a certain magnitude. With realistic observation duration, approximately an hour of exposure, the dimmest absolute magnitude cutoff corresponds to $m < -17.0$ or an apparent magnitude of $M \sim 24.0$. This yields a signal-to-noise (S/N) of around 5 for spectroscopic instruments such as the Keck II DEIMOS. We adopt this cut off as we advocate for securely identifying galaxy memberships of the cluster through the use of spectroscopic redshifts. Foreground galaxies can either be mistaken as the brightest cluster galaxy (BCG) or be a severe source of contamination in the luminosity map. This magnitude cutoff implies that we only include subhalos with at least several hundreds of stellar particles. We note that the richness of

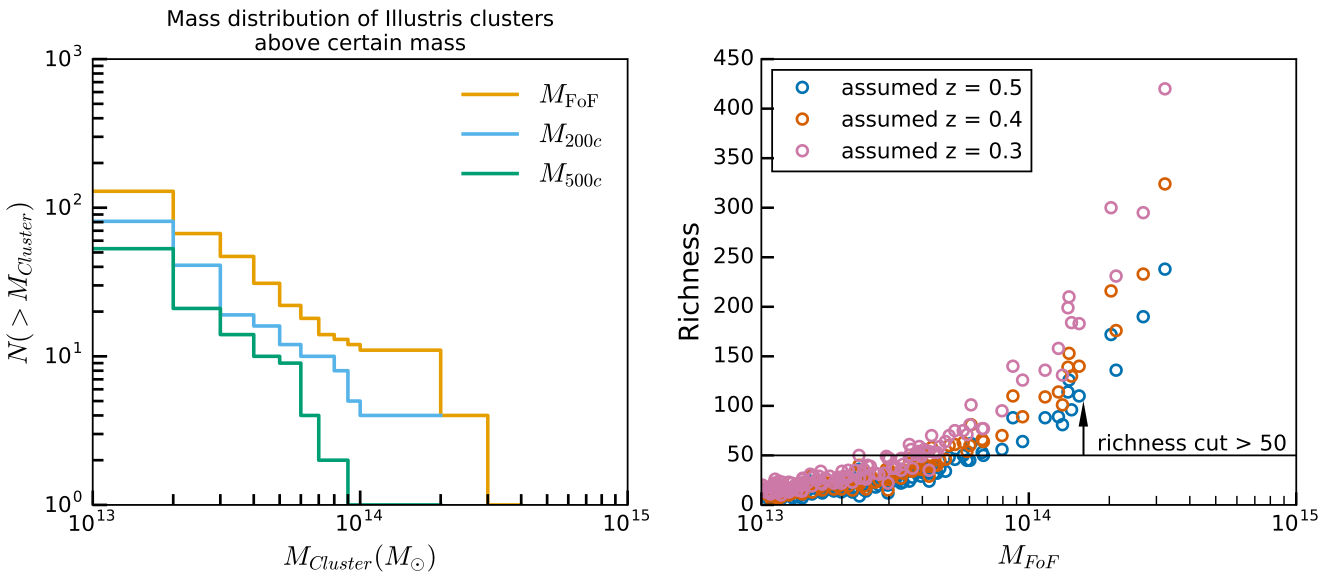

Figure 1. Left figure: Mass distribution of the group / cluster sized DM halos for different halo selection schemes, either selected by the Friends-of-Friends RockStar algorithm (M_{FoF}), or within the radius in which the average density is 200 or 500 times the critical density of the universe (M_{200c} and M_{500c} respectively). Discrepancies between the different measures of mass of the clusters indicate the presence of spatially separated substructures for the clusters. (See Fig ?)**Right figure:** Mass-richness relationship. We require clusters to have more than 50 member galaxies that are above observation limit, i.e. apparent $i \leq 24$ when we assume a cosmological redshift of $z = 0.4$, as shown by the richness cut.

a cluster is relatively more sensitive to the choice of limiting magnitude than the field of view for wide-view telescopes such as the Subaru suprime camera. Finally, for our final results, we only make use of galaxy clusters / groups that have at least 50 member galaxies after this magnitude cut.

2.3 Selection of the field of view

As a default output from the Illustris simulation, subhalos and particles of each galaxy cluster and group are identified by the Rockstar halo finder (CITATION). We make use of the member particle / 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. Furthermore, assuming a conservative line-of-sight (los) distance , i.e. cosmological redshift, of $z = 0.4$, the projected extent for most of the Illustris galaxy clusters, fits inside the field of view of telescopes, such as the Subaru Suprime Camera, which covers a physical area of ~ 9 Mpc $\times 7$ Mpc.

2.3.1 Spatial Projections

The center findings are all based on two-dimensional (2D) matter projections. In order represent the projection uncertainty, we sample the projections evenly by using HealPy (CITE), which is a Python wrapper for HealPix (CITE).

2.3.2 Galaxy weights

Not all galaxies are created equal. Some of them are more massive

One of the most common weighting scheme employed for cluster galaxy data is to weight by the luminosity in a particular band.

2.3.3 Cluster properties

2.4 Relaxedness of the clusters

On the relaxedness of the clusters. We provide several definitions of non-relaxedness to characterize whether the clusters underwent any recent merger activities. 1) These definitions of non-relaxedness. 2)

3 METHODS FOR SUMMARIZING THE SPATIAL DISTRIBUTION OF ??

There are many reasonable models for summarizing the overall spatial distribution of cluster components. Each method has different uncertainties. They fall under several categories, 1) mixture models, 2) wavelet or other basis expansions, and 3) non-parametric estimations such as a kernel density estimation.

Not only do the performance of the first two methods depend heavily on model parameters, the data fit also depend quite strongly on the underlying mixture model / wavelet basis. Often times, the most common mixture models and wavelet basis carry symmetry assumptions that may not be valid for galaxy clusters. This is because galaxy clusters have substructures over a wide range of length scales, from galaxy scales of hundreds of pc to fraction of a Mpc. models carry over symmetry assumptions that may not be valid.

Well known tradeoff Bias-variance tradeoff

Goal: to identify the "center" of the light distribution. Here the adopted tracers for the light distribution are the member galaxies of the cluster and groups.

We compare four ways to identify the light/galaxy centers:

(i) Centroids
(ii) KDE + peak finder
(iii) Shrinking aperture method
(iv) Brightest cluster galaxy (BCG)

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

Data	Selection strategy	Sensitivity	Relevant section
Field of view (FOV)	RockStar halo finder	comparable to FOV of the Subaru Suprime camera	
Observed filter	i-band	consistent over the redder r, i, z bands	
Richness of member galaxies	$i \leq 24$ and $z = 0.4$	sensitive to the assumed cosmological redshift of cluster and the assumed limiting magnitude of telescope	
Two-dimensional projections	even HealPix samples over half a sphere	discussed as results	

Table 2. Comparison between various methods for estimating one-point statistics of the galaxies of a cluster

Method	One-point statistic	Sensitivity to biases	Uncertainty	Relevant section
Centroid	2D spatial averages	High	Low	Not recommended
Shrinking aperture	proxy for density peak	High sensitivity to substructures	Medium	Not recommended
Peak finding from KDE	density peak	Lower sensitivity to substructures	Higher	Relatively computationally demanding bu
Brightest cluster galaxy		Sensitive to foreground contaminants	Canonical method	
Most bounded particle	bottom of gravitational potential well			unobservable but presented for compar

3.0.1 *Unweighted centroids*

We follow the usual definition of spatial centroid as

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_i \boldsymbol{x}_i. \tag{2}$$

While the weighted centroids are just:

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

for each spatial dimension and the weights w_i for the $i-th$ galaxy is described in section. Centroids can be biased 1) by subcomponents from merging activities yet do not provide explicit evidence for ongoing merger or accretion, 2) by the field of view.

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

We employed a KDE algorithm to infer a smooth density distribution of the galaxies while using smoothed cross-validation to obtain the optimal smoothing bandwidth matrices (H). Specifically, we made use of the KDE function in the statistical package ks (Duong) in the R statistical computing environment (R Core Team 2014). Cross validation eliminates free parameters in the KDE and minimizes the asymptotic mean-integrated squared error (AMISE) for a best fit to the data. After obtaining the KDE estimate, we employed a finite differencing algorithm to find the local maxima. We sorted the local maxima according to the KDE density at the maxima locations and identified the dominant peak.

Spatial location and the density of the subdominant peaks are also stored. We investigated if the presence of subdominant peaks are correlated with Δs.

3.0.3 *Shrinking aperture*

Another popular method among astronomers for finding the peak of a spatial distribution include what we call a shrinking aperture method. We test if the shrinking aperture method is able to recover the highest peak reliably. 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 consist of several subcomponents so the peak estimate can be easily biased by substructures. Furthermore, the convergence rate for this iterative algorithm is not analytical and is dependent on the data. We present the convergence criteria for reference. We note that different implementation may result in different performances.

Data: subhalo that satisfy cuts as a galaxy

initial˙center = mean(data_array)
dist_array = euclidean˙dist(initial˙center, data˙array)
apert = get_90th_percentile(dist_array)
while *(newCenterDist - oldCenterDist) / oldCenterDist* $\geq$ *2e-2* **do**
  new data array = old data array within apert
  newCenter = mean value of new data along each spatial dimension
end

Algorithm 1: Shrinking aperture algorithm

3.0.4 *Brightest Cluster Galaxies (BCG)*

3.1 Comparison of the methods from test data

In order to examine the performance 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.
Fig 1. one Normal mixture
Fig 2. one big normal mixture and one smaller normal mixture
Fig 3. three bridged normal mixtures
We compare the properties and performance of each of the methods for finding the peaks of the galaxy and dark matter, except the BCG since it does not rely on the cluster member population. The main factors that affect the performance of the methods depend heavily on statistical fluctuations of the drawn data. Namely, the performance of each method depends on: 1) the number of Gaussian mixture used, 2) the number of data points in each mixture

Due to the statistical nature of the data, it is not enough to just compare the performance from applying each method for one realization of the data. We also provide the 68% and the 95% confidence regions from the different methods for different Monte Carlo realizations.

The details and implementation can be found in our Bitbucket git repository.

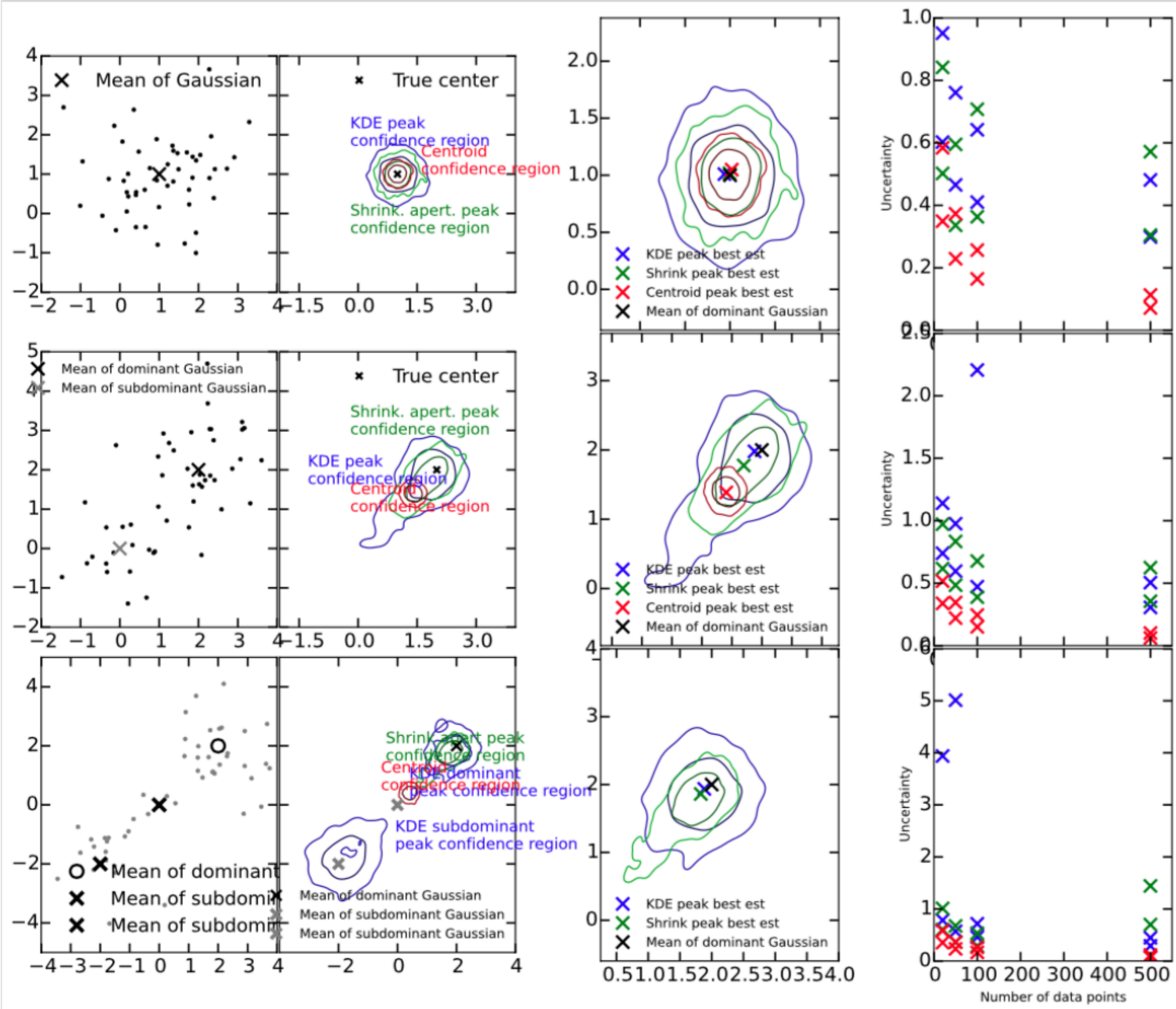

Figure 2.

4 SECTION II: DM AND THE LENSING KERNEL

To infer the 2D projected density, We reconstructed histograms of Since we have much higher resolution of dark matter, the choice of bin size have smaller impact on the results. The 2D histogram of the dark matter is analogous to a convergence map which is a product of a weak / strong lensing analysis.

4.1 Finding offsets

We computed the projected offsets between the galaxy density peaks inferred from the cross-validated KDE and the dark matter density peak as we have shown that this method gives us the least biased density peak estimates. The viewing angles of the projections are defined by an elevation angle ξ and an azimuthal angle ϕ.

5 RESULTS

5.1 Galaxy-DM Offset in Illustris

5.1.1 Projected offsets

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

6 DISCUSSION

6.1 Comparison to other simulations

6.2 Comparison to other observational studies

6.3 Galaxy-DM Offset in Merging Galaxy Clusters

7 SUMMARY

We showed that

- the peak finding method To-be-finalized for the density of cluster galaxies is the least biased due to substructures from our test data.
- all existing peak finding methods have non-negligible uncertainty due to the small number of data points. When dealing with small

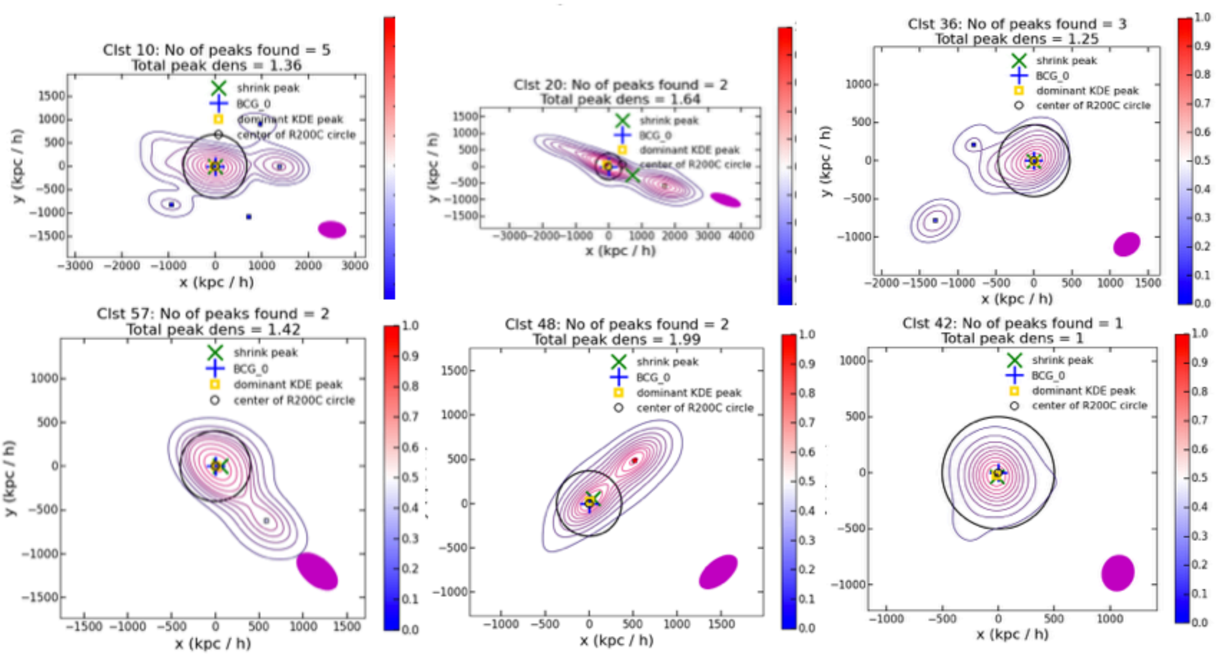

Figure 3.

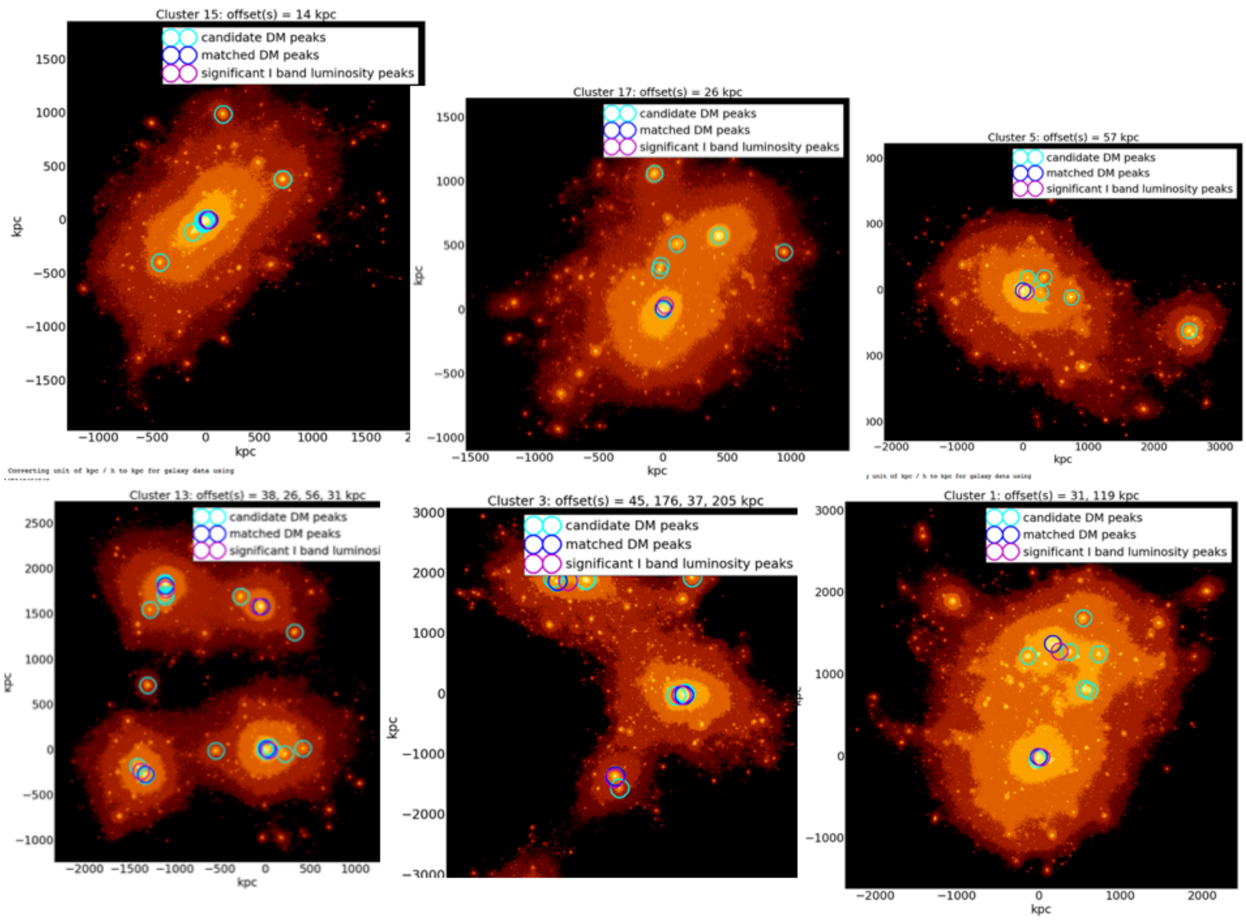

Figure 4.

number of cluster samples, the uncertainties of the peak locations should not be ignored.

8 ACKNOWLEDGEMENTS

K. Ng would like to thank Prof. D. Hogg and Prof. Thomas Lee for pointing out the ill-defined nature of modeling galaxy clusters.

APPENDIX A: THE PHYSICAL PROPERTIES OF A GALAXY CLUSTER

State variables are missing. A galaxy cluster is not a closed system.

APPENDIX A: KDE

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