DUST-OBSCURED GALAXY PROTOCLUSTER AND CLUSTER SURVEY (DOGPACS): IDENTIFYING LARGE-SCALE STRUCTURES 9−10 BILLION LIGHT-YEARS AWAY

A Dissertation IN
Physics
and
Electrical and Computer Engineering

Presented to the Faculty of the University
of Missouri–Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
RIPON SAHA

M.S., University of Missouri-Kansas City, Kansas City, Missouri, 2016
B.Sc., Bangladesh University of Engineering & Technology, Dhaka, Bangladesh, 2012

Kansas City, Missouri
2022

DUST-OBSCURED GALAXY PROTOCLUSTER AND CLUSTER SURVEY (DOGPACS): IDENTIFYING LARGE-SCALE STRUCTURES 9−10 BILLION LIGHT-YEARS AWAY

Ripon Saha, Candidate for the Doctor of Philosophy Degree

University of Missouri–Kansas City, 2022

ABSTRACT

Primordial density fluctuations have grown over time due to gravitational instability to form galaxies and, eventually, large-scale structures, such as clusters of galaxies. Galaxy clusters are the most massive collapsed structures in the universe. During cluster formation, the largest aggregation of gas, galaxies, and dark matter passes through an intermediate phase called the protocluster. Over the past decades, many studies have identified distant clusters and protoclusters due to advanced observational strategies. However, the protocluster-to-cluster transformation is still unclear, mainly due to the lack of large samples of early-stage clusters and late-stage protoclusters. Our research has identified a large selection of nearly 300 galaxy cluster candidates at redshift $1.3 < z < 1.8$ (9-10 billion light-years away) during the formation epoch of the galaxy clusters. This study leverages the fact that

high-z galaxy clusters and protoclusters exhibit enhanced star-formation and AGN activity in their cores. The candidates are identified using a sample of highly star-forming and/or AGN Ultra-Luminous Infrared Galaxies called the Dust-Obscured Galaxies (DOGs) as signposts in the Spitzer Deep Wide-Field Survey (SDWFS) in Boötes. A two-point correlation function analysis demonstrates that the sample has a mass scale of the galaxy clusters. Using a more multi-wavelength SDWFS catalog, this study has also uncovered a proto-supercluster structure at $z = 1.75$ ($\sim$10 billion light-years away). This proto-supercluster is a bound structure hosting dozens of protoclusters and clusters of galaxies, including the most massive galaxy cluster (IDCS J1426.5+3508) found to date at $z > 1.5$. Follow-up studies of this supercluster will provide a comprehensive picture of protocluster-to-cluster transition and the evolution of its constituent gas and galaxies. Finally, we develop and implement a novel machine learning technique to determine the photometric redshift (photo-z) of the galaxies using a decision tree-based architecture as part of the mission work for the upcoming *Euclid Space Mission* (2023). The photo-z results are comparable with many other template fitting- and machine learning-based methods.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Dust-Obscured Galaxy Protocluster and Cluster Survey (DOGPACS): Identifying Large-Scale Structures 9−10 Billion Light-Years Away" presented by Ripon Saha, candidate for the Doctor of Philosophy degree, and hereby certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Mark Brodwin, Ph.D., Committee Chair
Department of Physics & Astronomy

Daniel H. McIntosh, Ph.D.
Department of Physics & Astronomy

Paul Rulis, Ph.D.
Department of Physics & Astronomy

Masud Chowdhury, Ph.D., Inter-disciplinary Committee Chair
Department of Computer Science & Electrical Engineering

Faisal Khan, Ph.D.
Department of Computer Science & Electrical Engineering

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGEMENTS

This work is based on the data products from the now-defunct *Spitzer Space Telescope* which was operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. We acknowledge the use of the *Euclid*-like data products constructed by the OU-PHz unit of the upcoming *Euclid Space Mission* (2023). This study has extensively used the core python package for the astronomy community `Astropy` and `Scikit-Learn`. This work has been partially supported by funding from the UMKC SGS Research Grant received in the academic year 2019-2020. We are grateful to Daniel H. McIntosh, Bandon Decker, Benjamin Floyd, Gillen Brown, and Kameswara Bharadwaj Mantha for their valuable comments during the University of Missouri-Kansas City (UMKC) astronomy research meetings.

CHAPTER 1

GALAXY CLUSTER CANDIDATES SIGNPOSTED BY DUST-OBSCURED GALAXIES AT $1.3 \lesssim z \lesssim 1.8$

Unpublished: currently under review

Abstract

Over the past decade, high-redshift galaxy cluster studies have revealed a high concentration of star formation and AGN activity in cluster cores at $z \gtrsim 1.4$. This connection suggests that dusty, highly star-forming and/or active high-redshift galaxies, such as Ultra-luminous Infrared Galaxies (ULIRGs) or Dust-Obscured Galaxies (DOGs), might be employed to detect high redshift galaxy clusters. In order to test this hypothesis, we studied the surroundings of ≈ 2600 Dust-Obscured galaxies (DOGs) in the ~ 10 deg^2 *Spitzer* Deep, Wide-Field Survey (SDWFS) in Boötes. We developed and tested a pipeline to isolate and rank our detections, and using it have found 308 cluster candidates at $1.3 \lesssim z \lesssim 1.8$ above a 3σ detection significance. This pipeline has re-discovered nearly 60% of all the IRAC Shallow Cluster Survey (ISCS) clusters lying at $z \geq 1.3$ in Boötes, including IDCS J1426.5+3508 at $z = 1.75$, the most massive galaxy cluster known at $z > 1.5$. The two-point angular autocorrelation function of the cluster candidates was computed to measure their clustering strength. We derived the correlation length $r_o \approx 16\ h^{-1}$Mpc, which is consistent with the results from other infrared distant cluster search methods and

demonstrates conclusively that the sample has the mass scale of galaxy clusters. We computed the minimum dark matter halo mass consistent with the cluster candidate sample, $M_{200,min} \approx 2.34 \times 10^{13}\ M_{\odot}h^{-1}$, by comparing our detections to the halo mass functions (HMFs) from analytical prescriptions. These candidates will eventually grow by $z = 0$, occupying halos with, $M_{200,min} \approx 1.4 \times 10^{14}\ M_{\odot}h^{-1}$. We conclude that distant, highly star-forming galaxies selected in the mid-IR with *Spitzer* are sensitive signposts of the massive cluster and protocluster halos at the highest redshifts.

1.1 Introduction

As the largest collapsed structures, galaxy clusters facilitate the study of galaxies, gas, and dark matter in the deepest gravitational wells of the universe. Galaxy clusters started forming within redshift (z) $\gtrsim 2$ protoclusters at the intersection of filaments of the cosmic web [23]. Over the past few decades, advancements in multi-wavelength observations have permitted detection of distant, mature galaxy clusters [e.g. 49, 91, 111, 123, 128] during the early epoch of their formation, providing a 10 Gyr baseline for evolutionary studies. However, a conclusive evolutionary picture of the transformation of protoclusters into clusters via gravitational collapse and subsequent virialization is still unclear, primarily due to the lack of statistically complete early-stage cluster catalogs at $z \gtrsim 1.5$.

Although the central, hot, and ionized intracluster medium (ICM) shines up in X-ray wavelengths, only a few wide-field X-ray surveys have identified clusters at

$z \gtrsim 1.5$, such as the XDCP clusters [43], XMM-LSS distant cluster sample [122]. This paucity is largely due to X-ray surface-brightness dimming at such high redshifts. The other plausible reason is that the hot ICM may not have formed yet during the epoch of cluster formation. Only deep X-ray follow-up cluster studies with long exposure times have been successful to detect and characterize extremely massive objects at $z \sim 2$, such as IDCS J1426.5+3508 at $z = 1.75$ [15], SPT-Hiz cluster sample at $1.2 < z < 1.9$ [46, 85], XLSSC 122 at $z = 1.99$ [80], and SPT-CL J0459-4947 at $z = 1.71$ [79]. The *South Pole Telescope* has yielded a large sample of massive galaxy clusters ($> 10^{14} M_{\odot}$) out to $z \sim 1.5$ via measuring the Sunyaev-Zeldovich (SZ) decrement [12, 13, 61] — a distortion in the Cosmic Microwave Background (CMB) radiation caused by the intervening hot ICM electrons — which, in principle, does not depend on redshift [70]. Therefore, the ongoing SPT-3G survey [9] is predicted to detect dozens of massive clusters ($\sim 10^{14} M_{\odot}$) out to $z \sim 2$ as it is ~ 10x deeper than the initial SPT surveys.

Besides X-ray and SZ techniques, space telescopes, such as the *Spitzer Space Telescope* and the *Wide-field Infrared Survey Explorer* (WISE), have also detected many $z > 1$ galaxy clusters using stellar-mass-based methods in the infrared wavelengths [5, 39, 52, 89]. Due to cosmological K-correction, optical surveys fail to detect these clusters above $z \sim 1$ [84, 94]. The tracer technique detours the ICM- and galaxy-based high-z cluster finding algorithms. High-redshift galaxies, known as good cluster-tracers from previous surveys, are targeted to find the most massive structures around them. This technique leverages the fact that the high-z clusters

exhibit elevated star-formation (SF) and active galactic nuclei (AGN) activity in their cores compared to their low-z counterparts.

As has been long established [11, 36], most of the galaxy clusters in the local universe ($z \sim 0$) are swarms of bright-red, passively evolving early-type galaxies with little or no ongoing SF or AGN activity. Locally, the field-relative fraction of star-forming galaxies or AGNs declines as we move towards clusters' core regions [26, 62, 93, 101, 119]. Many studies have revealed a stark reversal of the clusters' local SF(AGN)-density trend at $z \gtrsim 1.4$ [6, 82, 109, 117]. Follow-up observations in the mid-and far-infrared (FIR) wavelengths have discovered a strikingly high abundance of dusty star-forming galaxies (DSFGs) and AGN-dominated galaxies in the galaxy cluster cores at $z \gtrsim 1.4$ [2, 19]. Indeed, the SPHerIC survey [81] identified many forming clusters surrounding far-infrared selected, extremely dusty star-forming galaxies at $1.3 < z < 3.0$. Several other studies have also detected many high-z clusters around radio-AGNs [45, 92, 97, 125].

In this study, we develop and validate a pipeline to identify distant clusters by using a sample of Dust-Obscured Galaxies (DOGs) as signposts [34] in Boötes. DOGs are optically faint, ultraluminous infrared galaxies (ULIRGs), powered by a combination of star formation and AGN activities at $1.5 \lesssim z \lesssim 2.5$. In a merger-driven scenario, DOGs are caught at the end stages of the mergers between two gas-rich disks [58, 90]. Such activities produce an enormous amount of dust that blocks the optical light and re-radiates the energy in the infrared part of the spectrum, making the DOGs extremely bright in the infrared wavelengths. As a feasibility

study, we explore the relationship first between the DOGs and a cluster sample from a previously known cluster survey, the IRAC Shallow Cluster Survey (ISCS) at $z > 1.3$ in Boötes. We present the DOG and ISCS sample in §2.2, and describe their relationship in §1.3. In §1.4, we demonstrate a method to identify new clusters using the DOGs as tracers. We characterize our detections in §1.5, and discuss the results in §1.6.

We use the following cosmological parameters throughout the paper: $\Omega_m = 0.3$, $\Omega_\Lambda = 0.70$, and $H_0 = 100h$ kms^{-1}Mpc^{-1}. A Hubble parameter $h = 0.7$ gives a physical scale of ~ 500 kpc/ in the redshift range, $z = 1.3 - 2.0$.

1.2 Data

1.2.1 The High Redshift DOG Sample

Over the past two decades, multi-wavelength observations have been carried out across the ~ 9 deg^2 Boötes field. The *Spitzer Space Telescope* has imaged the Boötes field at 24 μm using the onboard MIPS detector [Multiband Imaging Photometer for *Spitzer*, 105]. The NOAO Deep Wide-Field Survey [NDWFS, 65] is a deep optical survey in Boötes, carried out in the Kitt Peak National Observatory (KPNO). Using these observations, Dey et al. [34] identified a class of galaxies—a sample of 2603 optically faint, reddest 24 μm sources with $F_{24}/F_R \gtrsim 1000$—as Dust-Obscured galaxies (DOGs). The extreme red color, i.e., dust, is attributed to the ongoing star-formation and/or AGN activity in the DOGs. We adopted this sample to explore its relationship with high-redshift galaxy clusters for this study.

The absence of broad photometric features in the spectral energy distribution (SED) and inaccurate empirical SED templates prevent us from computing precise photometric redshifts for the DOGs. However, a subset of the DOGs ($\simeq 100$) has been spectroscopically confirmed via detection of its strong, narrow silicate absorption lines, PAH (Polycyclic Aromatic Hydrocarbon) features, or optical emission lines [33, 34, 59, 121]. These studies have revealed the Gaussian-like, broad spectroscopic redshift distribution of the DOGs with $\bar{z} \sim 2$ and $\sigma_z \sim 0.5$. We assume this distribution is valid for the entire DOG population throughout this study.

1.2.2 The IRAC Shallow Cluster Survey (ISCS)

The IRAC Shallow Cluster Survey [ISCS, 39] has unveiled 335 galaxy clusters at $0 < z < 2$ in Boötes. The ISCS clusters were detected via a photometric redshift-based cluster-search pipeline using the IRAC Shallow Survey [ISS, 38] data. ISS has mapped Boötes in all four *Spitzer* IRAC [Infrared Array Camera, 44] bands. Photometric redshifts were computed by matching the IRAC 4.5 μm-selected ISS sources to the NDWFS data for the optical (B_wRI) photometry [16]. Subsequently, Eisenhardt et al. [39] identified the ISCS clusters via searching for overdensities of the ISS sources in narrow photometric redshift slices. As the bulk of the DOGs is at high redshift ($z \sim 2.0 \pm 0.5$), we isolated and used 50 ISCS clusters at $z \geq 1.3$ to test any spatial connection with the DOGs. Of these, 10 ISCS clusters have been confirmed spectroscopically at $1.3 \lesssim z \lesssim 1.5$ [19, 110, 129]

1.3 Relationship Between the DOGs and the ISCS Clusters

Previous studies have identified the ISCS clusters and the DOGs with two different detectors: 4.5 μm and 24 μm, respectively [34, 39]. Before testing any physical link between them, we must construct the cluster and the DOG sample from the parts of the sky with perfect imaging in both wavelengths. We chose the mask image that was used to identify the ISCS clusters [14, 16]. The areas with bright stars and various CCD artifacts, such as diffraction spikes, saturation tails from bright stars, bad image columns, were rejected. The total area with perfect imaging (unmasked) is 7.23 deg^2. All 50 high-z ISCS clusters are in the unmasked area by our choice. Approximately 87% of the DOGs (2259) are useful for probing the physical association, as only these DOGs lie in the unmasked regions.

1.3.1 Surface Number Density of the DOGs

We computed the projected surface density of these DOGs around the high-z ISCS clusters as a function of distance from the cluster centers and compared that with the average surface density of the DOGs (0.087 arcmin^{-2}). The surface density was measured up to 6arcmin ($\approx$ 3 Mpc) away from the cluster centers. The circular patch of 6 arcmin radius around each cluster center was divided into multiple radial slices. The width of the radial slices is not equal, chosen to achieve a statistically significant signal-to-noise ratio in each of the radial bins. We carried out 5000 bootstrap simulations to estimate the uncertainties in the measurements. For each simulation, a sample of 2259 DOGs was re-sampled with replacement. The various

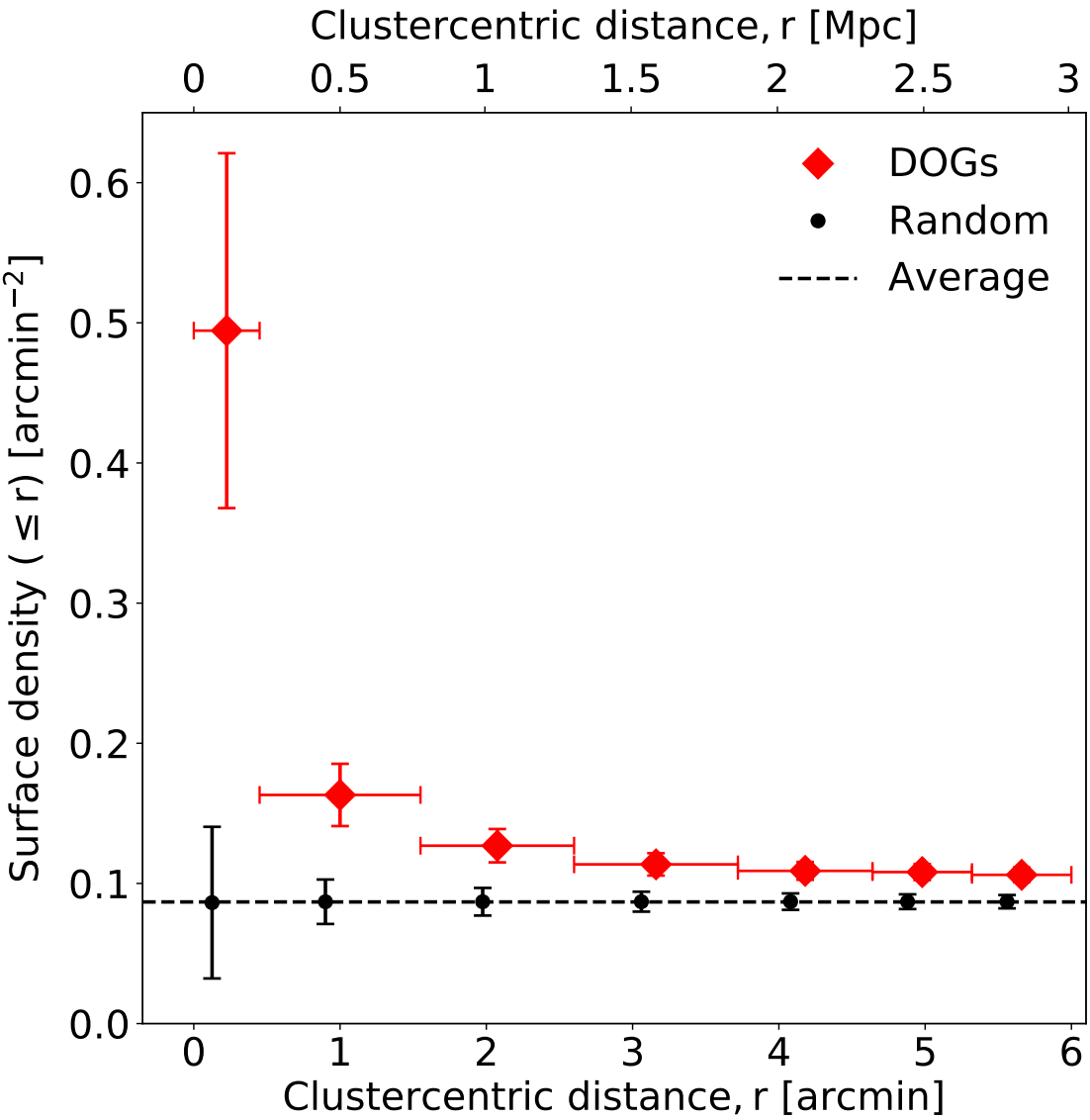

Figure 1: The projected cumulative surface density of the DOGs around the $z \geq 1.3$ ISCS galaxy clusters (red diamonds). The horizontal bars represent the radial bin sizes. The vertical error bars depict the 1σ uncertainties on the measurements. The dashed horizontal line represents the mean density of the DOGs (0.087 arcmin^{-2}) in Boötes. The black circles represent the surface density of the randomly distributed objects around the clusters, which is consistent with the average value across all the radial bins. Circular symbols are shown with a slight offset for clarity. The innermost region of the ISCS clusters has a clear excess density of the DOGs.

Table 1. Annular surface number density of the DOGs

r (arcmin)	Annular Surface Density (arcmin^{-2})	Significance (in σ)
r $\leq$ 0.45	0.494±0.127	3.21
0.45 $<$ r $\leq$ 1.55	0.132±0.021	2.15
1.55 $<$ r $\leq$ 2.60	0.106±0.014	1.37
2.60 $<$ r $\leq$ 3.72	0.100±0.011	1.20
3.72 $<$ r $\leq$ 4.64	0.100±0.010	1.32
4.64 $<$ r $\leq$ 5.32	0.105±0.011	1.66
5.32 $<$ r $\leq$ 6.00	0.099±0.010	1.22

Note. — r is the clustercentric distance. The detection significance is defined as the field-subtracted signal to noise ratio.

counts of the DOGs in different simulations follow a Gaussian distribution. The standard deviation of this distribution yields uncertainty on the incidence rate of the DOGs. We also computed the surface number density of a randomly distributed Poisson population around the cluster sample. We generated 5000 random catalogs. Each catalog contained 2259 (equal to the number of DOGs) randomly generated right ascension and declination pairs. The average count of the random points and the error on the random expectation were used to compute the surface density for the Poisson catalog.

Figure 1 shows the surface number density of the DOGs around the ISCS clusters as a function of the clustercentric distance. The density profile is indicative of a strong spatial correlation between DOGs and their host clusters. The surface

density of the DOGs is remarkably high in the cores, particularly within ~ 0.5 arcmin of the galaxy cluster centers. Indeed, DOGs are overdense by a factor of 5.6 ± 1.4 in the central region of the clusters relative to their average incidence rate in Boötes. The surface density approaches the field value as we move away from the cluster cores. This trend strongly suggests that the DOGs prefer to reside often within ~ 250 kpc of the cluster center. The detailed results are listed in Table 1.3.1. On the contrary, the density of the random points around the cluster sample remains consistent with the average occurrence rate of the DOG sample. This indicates that DOGs are not randomly located. The high detection significance ($\sim 3\sigma$) close to the cluster cores indicates a statistically significant result. Remarkably, even at the outermost annulus, the overdensity signal remains statistically significant at $\gtrsim 1\sigma$ level. Perhaps, the dense high-z cluster environment plays a favorable role in hosting the extremely dusty ULIRGs, such as DOGs at $z \gtrsim 1.3$.

This radius-density relationship is not dependent on the mid-infrared brightness (24 μm flux) or dustiness (R$-$[24] color) of the DOGs. We divided the DOG sample into three flux (color) bins, chosen to have roughly equal numbers of DOGs in each bin. We observed similar overdensity trends for all the flux or color bins. The following section presents a spectroscopic observation of a DOG near an ISCS cluster core.

1.3.2 DOGs in ISCS J1433.9+3430: A Spectroscopic Evidence

We report a spectroscopically evident physical connection between a DOG and a cluster ISCS J1433.9+3430 [α=14:33:57.82 & δ=34:30:40.86, 39]. The Deep Imaging Multi-Object Spectrograph [DEIMOS, 42] on the Keck II telescope obtained the spectra of the objects in the vicinity of ISCS J1433.9+3430. Two DEIMOS slitmasks obtained high-quality spectra of 134 sources. Of which, 38 *line-of-sight* galaxies were identified within 5 arcmin ($\approx$ 2.5 Mpc) of the cluster center. Owing to the strong [OII] emission lines, eight galaxies were detected at spectroscopic redshift, $z_{spec} = 1.419^{+0.014}_{-0.013}$. This detection provides spectroscopic confirmation of the cluster ISCS J1433.9+3430. The left panel in Figure 2 shows the redshift distribution of the Keck/DEIMOS slitmask observation (blue histogram) near the cluster, which has a peak around $z = 1.4$. Spectroscopic redshifts of the member galaxies are provided in Table 2.

We also identified three DOGs within 5 arcmin of the cluster center. One of these DOGs (α=14:33:53.67 and δ=34:31:56.67) has a spectroscopic redshift z_{spec} = 1.406 [for detailed spectroscopy, see 34]. The solid green line in Figure 2 (left panel) shows the redshift of this DOG. No redshift information is available for two other *line-of-sight* DOGs near the cluster center. The right panel in Figure 2 shows the IRAC 4.5 μm map of ISCS J1433.9+3430. The DOG with spectroscopic redshift is located only $\approx$ 90 away from the cluster center along with two spectroscopic cluster members. This proximity to the cluster center demonstrates a close association

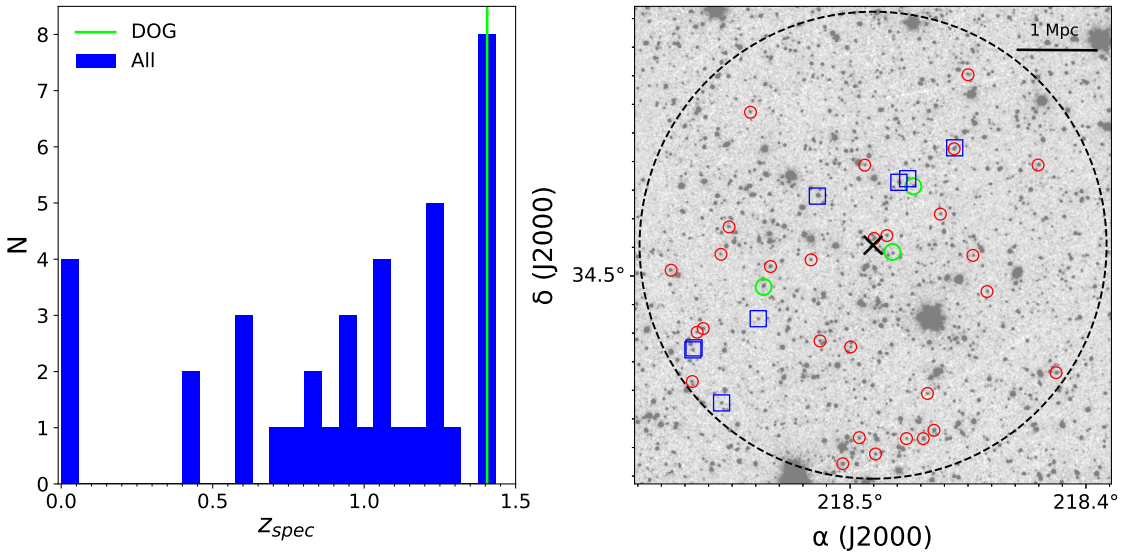

Figure 2: *Left*: The blue histogram depicts the Keck/DEIMOS spectroscopic redshift distribution of all the *line-of-sight* galaxies within 5 arcmin of the galaxy cluster, ISCS J1433.9+3430 [39]. Eight (8) of these objects are located at $z_{spec} = 1.419^{+0.014}_{-0.013}$ and are deemed as cluster members. The solid green line represents the spectroscopic redshift of a DOG at $z_{spec} = 1.406$, located within ≈ 90 from the cluster center [34]. *Right*: IRAC 4.5 μm gray-scale image of the cluster ISCS J1433.9+3430. The ~ 10 arcmin $\times$ 10($\approx$ 5 Mpc $\times$ 5 Mpc) image is centered on the cluster, and the black dashed circle represents a ~ 5 arcmin (≈ 2.5 Mpc) radius around the cluster center (black cross). North is up, and east is to the left. The blue squares represent the locations of the eight spectroscopic cluster members. Three DOGs (green circles) are identified within 5 arcmin from the cluster center. The northern DOG is spectroscopically confirmed at $z_{spec} = 1.406$. Red circles mark the locations of the potential cluster members, i.e., galaxies with photometric redshifts at $1.35 < z < 1.45$.

Table 2. Spectroscopic members of the galaxy cluster ISCS J1433.9+3430

R.A. (J2000)	Decl. (J2000)	z_{spec}	Spectral feature
218.480154	34.533414	1.433	[OII]
218.514708	34.528389	1.432	[OII]
218.476417	34.534689	1.429	[OII]
218.539442	34.485372	1.424	[OII]
218.456542	34.545583	1.418	[OII]
218.554738	34.455833	1.407	[OII]
218.567067	34.474306	1.406	[OII]
218.566583	34.475	1.406	[OII]

Note. — The table entries are sorted according to their spectroscopic redshifts in a descending order

between a DOG and a high-z galaxy cluster. Together with the DOGs (green circles), the spectroscopic members (blue squares) trace an elongated structure along the southeast−northwest direction. Figure 2 also shows the location of the galaxies with photometric redshifts [16] at $1.35 < z < 1.45$ (red circles), which possibly demonstrates the true extent of the cluster. This spectroscopic evidence, along with the conclusive demonstration of the excess surface density of the DOGs in the high-z ISCS clusters (§1.3.1), makes it reasonable to test the DOGs as tracers of high-z structures. The following section proposes and validates a new method to identify galaxy clusters using DOGs as signposts at $z > 1.3$.

1.4 Dust-Obscured Galaxy Protocluster And Cluster Survey (DOGPaCS)

1.4.1 The Spitzer Deep, Wide-Field Survey

For the cluster search, we adopted a deeper survey than ISS in Boötes. Motivated by the outcomes of the ISS survey, such as detection of many high-z clusters [14, 39, 110], active galaxies [57, 113], and galaxy evolution studies [20, 30], the Boötes field was remapped in the *Spitzer* IRAC bands (3.6 μm, 4.5 μm, 5.8 μm, and 8.0 μm) three more times to the same depth of the original ISS survey. The campaign is known as the Spitzer Deep, Wide-Field Survey [SDWFS, 3]. For the purpose of this paper, we use a 4.5 μm-selected, 4 aperture catalog of $\sim 4 \times 10^5$ SDWFS sources to identify high-z clusters of galaxies. SDWFS reaches a 5σ depth of 18.84 mag (Vega) at 4.5 μm, which is $\sim$ 0.7 mag deeper than the ISS survey. Approximately 90% of the SDWFS sources are also detected in the NDWFS optical (BwRI) bands. This study selects distant galaxy cluster candidates based on the overdensity of high-z SDWFS sources traced by DOGs, and the search will be referred to as the DOGPaCS survey hereafter.

1.4.2 The Selection of High-z Galaxies

The search for high-z SDWFS sources began by removing the stars from the catalog. The SDWFS sources with SExtractor CLASS_STAR parameter value greater than 0.95 were deemed stars ($\approx 0.5\%$) and were removed from this analysis. We used the IRAC color, $[3.6] - [4.5]$ to select the SDWFS galaxies at the high

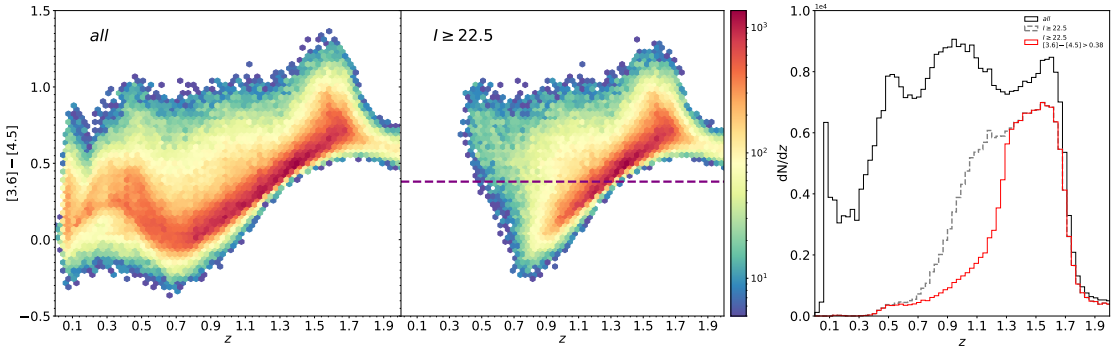

Figure 3: *Left*: The $[3.6]-[4.5]$ colors are shown as a function of the photometric redshift for all the SDWFS sources. The bulk of the sources has an increasingly red color at $z > 1$. The color appears similarly red for a small subset of galaxies at $z \sim 0.3$. *Middle*: A color cut (purple dashed line) effectively removes the bulk of the foreground contamination. A small number of the optically bright low-z galaxies (around $z \sim 0.3$) stand out against the cut. A magnitude cut, $I > 22$, finally isolates the galaxies at $z \gtrsim 1$. *Right*: Histogram (solid black line) showing the photometric redshift distribution for all the SDWFS sources. A small subset of galaxies at $z \sim 0.3$ contaminates the sample when only the color criterion is used (dashed black line). Finally, combining the color and magnitude criteria yields a sample with most of the galaxies at $z > 1.3$.

redshifts. This selection works because of the broad, rest-frame 1.6 μm emission from the in situ stellar population of the galaxies at $z \gtrsim 1.3$.

The primary source of opacity in the stellar atmospheres is the H^- ions. The opacity reaches a minimum level at around 1.6 μm [68]. This opacity leads to an emission bump at 1.6μm in nearly all stellar spectra, followed by a Rayleigh-Jeans tail. At $z \gtrsim 1$, the 1.6 μm peak shifts into the IRAC bands, and thus the galaxies have increasingly red $[3.6]-[4.5]$ color with increasing redshift until the color reaches a plateau at $z \sim 2$. The $[3.6]-[4.5]$ color appears bluer at $z < 1$ as it traces the Rayleigh-Jeans tail of the spectra at the low redshifts. Over the past decade, many studies have demonstrated the efficacy of using the $[3.6]-[4.5]$ color in isolating galaxies at $z > 1$ [52, 81, 89, 95, 104, 125]. Following these studies, we applied the color criterion, $[3.6]-[4.5] > 0.367$ (Vega), to separate the high-redshift SDWFS sources. Galaxies with signal-to-noise ratio, S/N ≥ 10 in both IRAC bands (3.6 μm and 4.5 μm.) were only considered. The S/N criterion confirms a sample of IRAC sources with large detection significance and small color errors.

We also computed the photometric redshifts (photo-z) of the SDWFS sources with 7-band photometry (KPNO BwRI and IRAC 3.6 μm, 4.5 μm, 5.8 μm, and 8.0 μm) using the method described in Brodwin et al. [16] with an accuracy, $\sigma_z = 0.10(1+z)$ at $z > 1$. Figure 3 shows the density map of the IRAC $[3.6]-[4.5]$ color distribution as a function of the photometric redshifts. The IRAC color $[3.6]-[4.5] > 0.367$ (Vega) selection identified the bulk of the high-redshift objects. The primary contaminants are infrared-bright star-forming galaxies at $z \sim 0.3$ that have a similar

red color, where the red color stems from the warm dust produced by starburst activity [63, 95]. To remove these galaxies from our sample, we applied an I-band magnitude cut, I $>$ 22 mag (Vega). Galaxies' optical emission has a very positive K-correction, for which the flux density decreases drastically with increasing redshift [22]. Hence, the I-band magnitude selection removed the foreground, low-z galaxies from the sample with a minimal effect on the high-z sources. Finally, combining the I-band magnitude and IRAC color selection yielded a sample of 83555 SDWFS galaxies at $z \gtrsim 1.3$, demonstrating the success in identifying high-z galaxies without photometric redshifts. Hereafter we will refer to these galaxies with $[3.6] - [4.5] > 0.367$ and I $>$ 22 as the IRAC high-z sample. Nearly 80% of the objects in this sample are at $z > 1.3$, and less than 2% of the objects are at $z < 1$.

1.4.3 Detection of the Densest Regions

We studied the surroundings of all the 2603 DOGs (DOGPaCS fields hereafter) in our sample from Dey et al. [34]. We began by counting the IRAC high-z galaxies —a counts-in-cell analysis—around the DOGs [45, 81]. This count was defined as the richness value (λ). Different circular aperture radii (r) were tested to begin the search. Smaller values of r than 1 arcmin are too small to contain any useful signal and do not represent galaxy clusters' R_{500} value at $z \sim 1.5$ [124]. Larger values of r than 1 arcmin often introduce many interlopers. Therefore, we adopted 1 arcmin as the search radius, corresponding to $\sim$ 500 kpc at $1 < z < 2$. Although DOGs prefer to live in rich environments, they are not necessarily the

center of the overdense regions. To solve this problem, we computed the centroids (center-of-mass) of the smoothed, 4.5 μm flux-weighted distribution of the IRAC high-z galaxies within 1 arcmin of the DOGs. These centroids were new centers for dropping the search aperture, and, therefore, we computed the richness around them. We continued iterating until the centers had converged or the richness had decreased. Thus, the iterative pipeline identified the largest collection of IRAC high-z galaxies, i.e., maximum richness value near the DOGs.

The iterative process normally converged in $\sim$ 97% of the fields. We ended the iteration early in $\sim$ 3% cases because of the decrement of the galaxy count. Two of the DOGs in the sample do not have any IRAC high-z galaxies around them and are removed from the analysis. Figure 4 illustrates an example of the final step of the iterative method to find the largest assembly of IRAC high-z galaxies near a DOG (α = 14: 32: 37.33 & δ = 34: 35: 58.21). The background colormap is a Gaussian-smoothed distribution of the IRAC high-z galaxies, where the bluest color marks the densest region on the map. This map was used to compute the centroid of the distribution. Starting by counting the galaxies around the DOG, the pipeline ultimately identified the densest region $\approx$ 40 away from the initial signpost in this example.

However, a DOG's association with its nearest densest region is not always unique. Two or more closely located DOGs can trace the same largest aggregation of the IRAC high-z galaxies. We considered two assemblies of DOG-signposted-population of high-z galaxies, i.e., DOGPaCS fields same based on two criteria: (i)

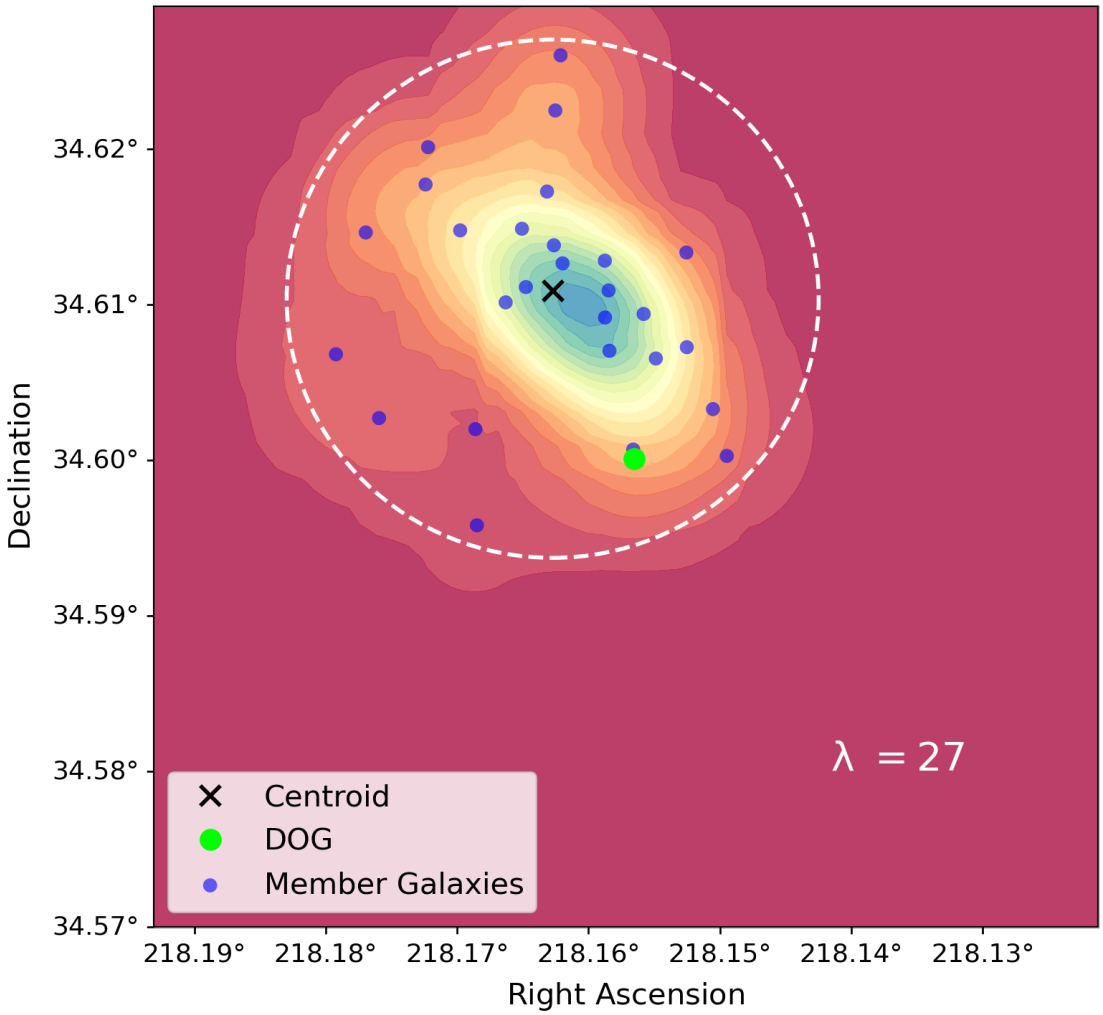

Figure 4: An example of a smoothed colormap of the flux−weighted spatial distribution of IRAC high-z galaxies in a 1 arcmin aperture near a DOG. The locations of the galaxies (blue circles), centroid (black cross), and DOG (green circle) are overplotted. The circles are not drawn to scale and hence, do not represent the actual sizes. The image shows the final step of the iterative pipeline after convergence, where the centroid of the densest area (richness, $\lambda = 27$) is found ~ 40 from the initial signpost (DOG). The dashed white circle represents the 1 search radius to find high-z galaxies within.

the projected separation between the centroids is smaller than the search radius (1); (ii) the difference between the mean photometric redshifts of the galaxies in two population is consistent with zero. For overlapping DOGPaCS fields, we computed the centroid of the galaxies in the regions combined and the richness by counting the galaxies within 1 arcmin of the new centroid. Finally, we identified 2065 DOGPaCS fields of IRAC high-z galaxies uniquely associated with one or more tracer DOGs. Of these, $\approx 20\%$ (429) fields are signposted by two or more DOGs, and the rest of the fields have only a single incidence of DOGs.

1.4.4 Cluster Candidacy

The DOGPaCS fields are not immune to data-quality problems. We used an SDWFS mask to compute the fraction of unsaturated—free from imaging issues—pixels in the circular apertures. This fraction is defined as the good pixel fraction (GPF). We scaled up the richness value in each field by its GPF value to account for the missing number of galaxies due to imperfect imaging. Fields with GPF values greater than 80% are only considered in this analysis to reduce the number of noisy detections. This threshold value retained 2021 DOGPaCS regions to search for galaxy clusters while removing only $\sim 3\%$ of the total regions.

The filled histogram (orange) in Figure 5 (left panel) illustrates the probability distribution of the up-scaled richness values (λ) for all the fields with GPF greater than the threshold. We compared this profile with that of a randomly distributed population. We dropped 1000 times more random circular apertures

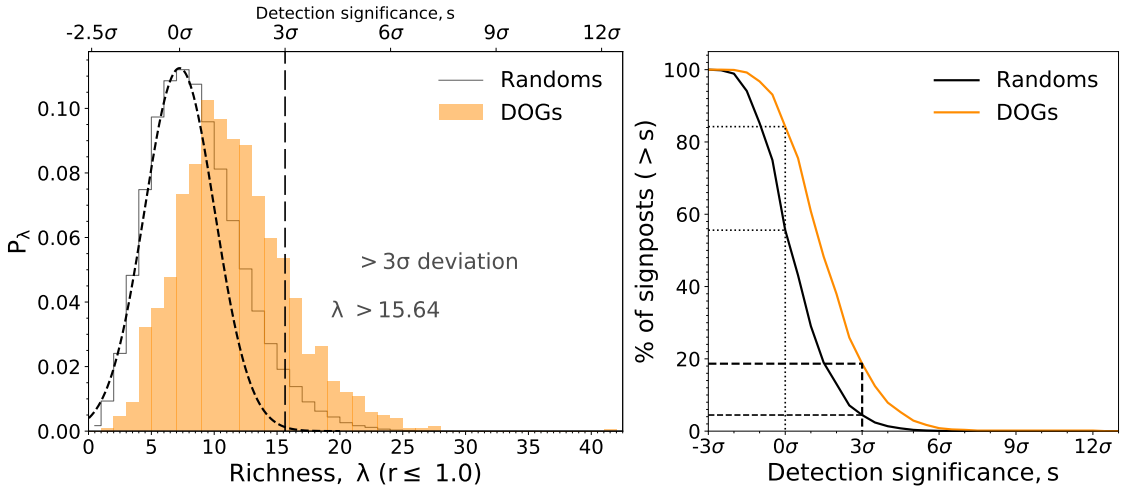

Figure 5: *Left*: Histogram showing the richness (λ) distribution for the DOGPaCS fields in orange. The corresponding distribution for the randoms is shown in a solid black line. The dashed black line is the best-fit Gaussian to the random profile. The standard deviation of the Gaussian is represented by σ. The top x-axis depicts the richness values in the units of detection significance (s). *Right*: The orange line shows the percentages of the DOGs residing in the DOGPaCS fields with significance $>$ s. The black line represents the corresponding profile for the randoms. The dashed line compares the fraction of signposts residing in regions with more than 3σ significance. The dotted line compares the fraction of signposts residing in regions with more than the average number of IRAC high-z galaxies.

($\sim 2.5 \times 10^5$) of 1 arcmin than the total number of DOGs across the SDWFS area for a robust computation. The histogram of the richness values, i.e., up-scaled number of IRAC high-z galaxies, around the randoms exhibit a Gaussian profile similar to the DOGs'. A skewed richness distribution for the DOGs with an evident, large offset from the random profile indicates that the DOGs, in general, prefer to live in richer environments. Of course, the random profile is also asymmetric. A tiny fraction of the random points coincide with the dense regions with richness, $\lambda > 15$ by chance. However, as demonstrated in the left panel of Figure 5, the chances of detecting such systems by dropping randoms are significantly lower than using the DOGs as signposts. To obtain the true richness profile in the absence of clustering, we fit a Gaussian to the lower richness half of the distribution for the randoms. From the best-fit Gaussian, we obtained the average number of IRAC high-z galaxies around the randoms, $\mu \approx 7.24$ and the standard deviation, $\sigma \approx 2.8$. Finally, we define DOGPaCS fields with richness, $\lambda \geq \mu + 3\sigma = 15.64$ as DOG-signposted overdensities, i.e., cluster candidates with detection significance, s $\geq 3\sigma$.

We identified 308 (125, 45) DOGPaCS cluster candidates with $\geq 3\sigma$ (4σ, 5σ) significance. Nearly 20%, 8%, and 3% of the total DOGs reside in $\geq 3\sigma$, 4σ, and 5σ candidates. For the random signposts, only 4.5%, 1.4%, and 0.3% of the fields are richer than $\mu + 3\sigma$, $\mu + 4\sigma$, $\mu + 5\sigma$, respectively. Figure 5 (right panel) shows the number of the signposts as a function of the overdensity significance of the fields in which they reside. Most of the DOGs (84%) reside in the overdense regions, whereas a large fraction of the random points (44%) coincide with the underdense

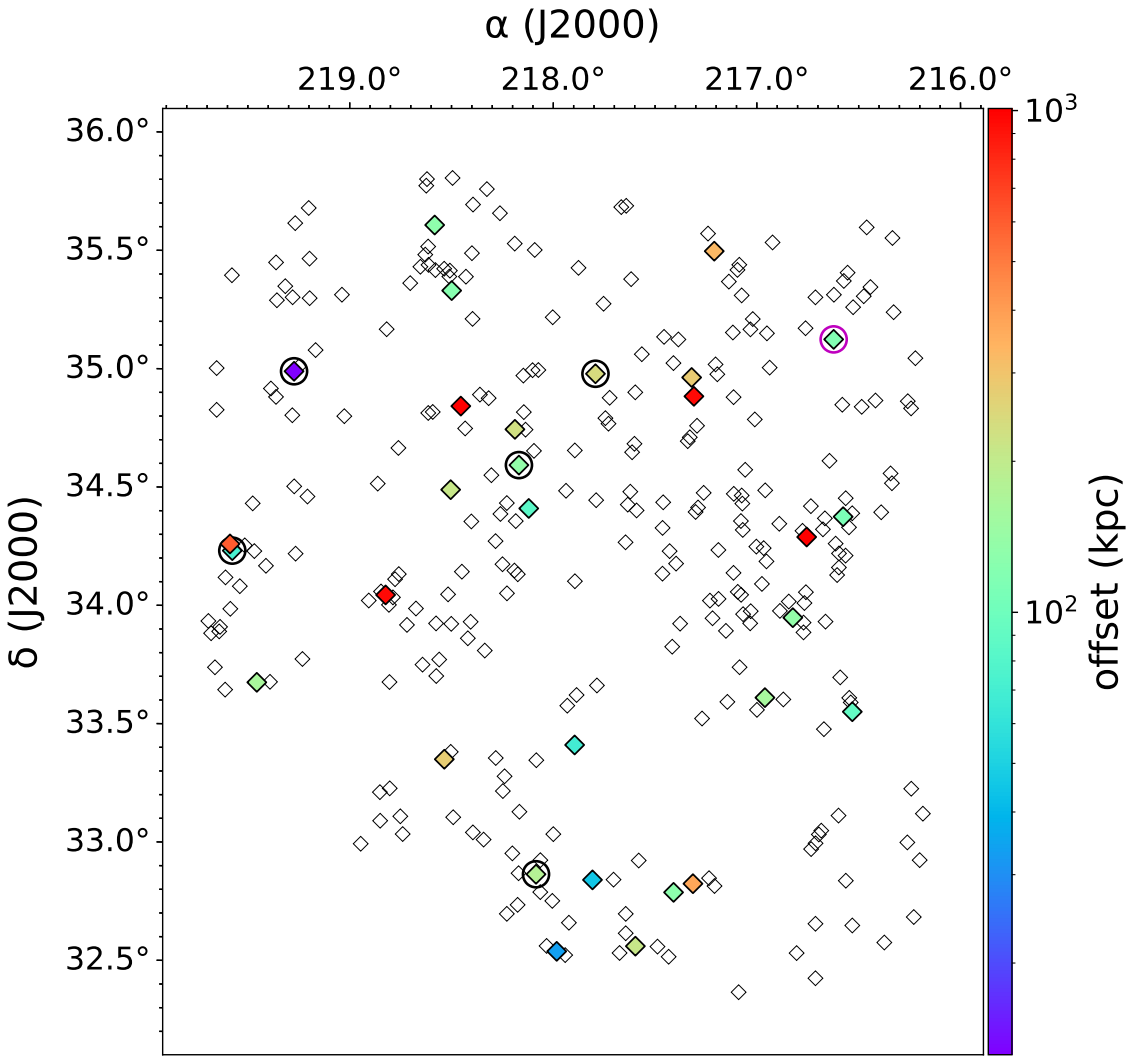

Figure 6: Projected distribution of the 308 DOGPaCS cluster candidates in Boötes at $1.3 \lesssim z \lesssim 1.8$. North is up, and east is to the left. Each data point represents the centroid of a cluster candidate in the projected sky. The data points are not drawn to scale. The filled diamonds are the re-discovered clusters previously identified by the preceding surveys at $z \geq 1.3$. The colorbar shows the matching offset in the units of kiloparsecs assuming $h = 0.7$. Filled diamonds with black circles around them represent the re-discovered five ISCS clusters that are spectroscopically confirmed at $z_{spec} > 1.4$. Please see § 1.6.2 for the details. The magenta circle marks the location of the spectroscopically confirmed IDCS 1426.5+3508 cluster at $z_{spec} = 1.75$, which is re-discovered by this study with a matching offset of ~ 125 kpc only (see § 1.6.3).

regions. Altogether, DOGs are significantly better cluster (overdensity) tracers. Figure 6 depicts the distribution of the DOGPaCS cluster candidates in Boötes. The cluster candidates are listed in Table 1.4.4. In terms of projected separation, $\sim 17\%$ of the total DOGs trace $\geq 3\sigma$ cluster candidates within 1 arcmin ($\sim$500kpc). The $\geq 3\sigma$ candidates were traced more than 2 arcmin away from their signposting DOGs only in four cases. Other biased tracer methods, such as the SPHerIC survey [81, $1.3 < z < 3$], used sub-millimeter-selected, extremely star-forming galaxies as signposts and detected clusters of IRAC sources in 27% of its total fields. The CARLA survey [126, $1.3 < z < 3.2$] identified clusters of IRAC sources within 1 arcmin of the 37% of the total fields around a sample radio-loud AGNs. However, these methods adopted different tracers, different galaxy surveys, and thus may not be suitable for direct comparison.

1.5 Angular Clustering of the DOGPaCS Candidates

Galaxy clusters are strongly correlated non-random structures. The two-point angular (auto)correlation function $\omega(\theta)$ is related to the excess probability of finding two clusters separated by a distance (θ) with respect to the random expectation [100]. Using the estimator proposed in Landy & Szalay [72], we computed the two-point angular correlation function (ACF) of the DOGPaCS cluster candidates ($\geq 3\sigma$) at angular scales $\theta \lesssim 1$.

$$\omega(\theta) = \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD}{RR} - \frac{N_R-1}{N_D}\frac{DR}{RR} + 1, \tag{1.1}$$

Table 3. List of the Top 20 DOGPaCS Cluster Candidates

Cluster Candidate	R.A. (J2000)	Decl. (J2000)	z_{phot}	Richness (λ)	Significance (s)
DOGPaCS J1438.1+3414	219.537297	34.23646	1.43	42.0	12.4
DOGPaCS J1436.1+3519	219.026411	35.324561	1.47	27.91	7.4
DOGPaCS J1432.6+3436	218.162406	34.6097	1.38	27.0	7.1
DOGPaCS J1438.2+3415	219.548844	34.264673	1.49	27.0	7.1
DOGPaCS J1428.4+3526	217.101252	35.433368	1.57	26.39	6.8
DOGPaCS J1437.5+3455	219.364504	34.924186	1.65	26.0	6.7
DOGPaCS J1427.7+3532	216.929074	35.548303	1.47	25.12	6.4
DOGPaCS J1428.3+3519	217.081704	35.325509	1.69	25.0	6.3
DOGPaCS J1432.9+3313	218.232952	33.231106	1.43	24.86	6.3
DOGPaCS J1428.6+3523	217.143513	35.384156	1.45	24.48	6.2
DOGPaCS J1437.0+3537	219.258909	35.62452	1.56	24.2	6.1
DOGPaCS J1438.4+3344	219.610026	33.742217	1.42	24.18	6.1
DOGPaCS J1431.2+3251	217.807716	32.856267	1.48	24.04	6.0
DOGPaCS J1434.6+3400	218.653091	34.000905	1.56	24.0	6.0
DOGPaCS J1432.2+3248	218.054323	32.804364	1.64	24.0	6.0
DOGPaCS J1436.8+3541	219.194015	35.689254	1.59	24.0	6.0
DOGPaCS J1428.9+3530	217.213725	35.513172	1.47	23.94	6.0
DOGPaCS J1429.2+3425	217.299411	34.429827	1.35	23.68	5.9
DOGPaCS J1425.8+3452	216.438202	34.875614	1.49	23.48	5.8
DOGPaCS J1435.6+3300	218.904839	33.004219	1.64	23.42	5.8

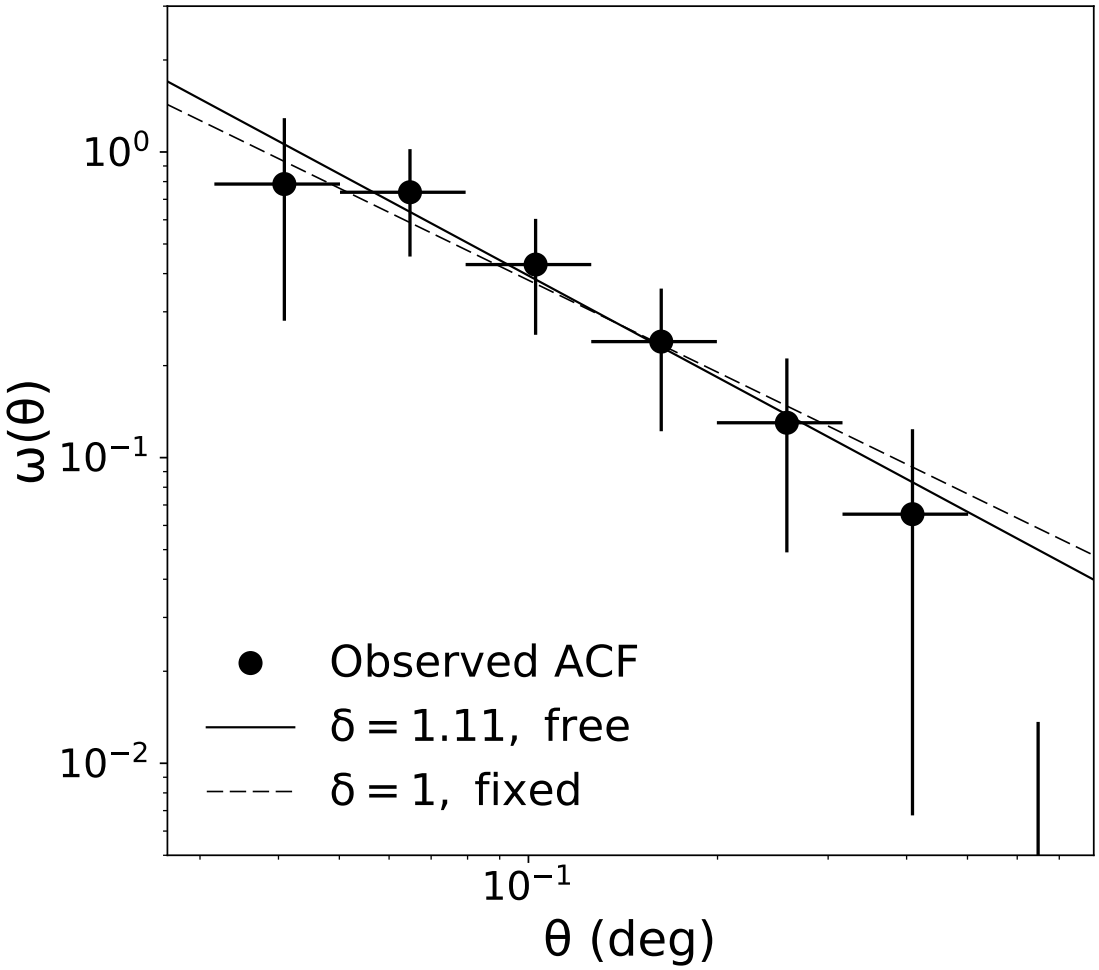

Figure 7: The data points represent the two-point angular correlation function (ACF) of the $\geq 3\sigma$ cluster candidates in the bins of angular separations. The horizontal error bars represent the width of the angular bins. The errors on the ACF measurements are computed using bootstrapping. The solid line is the best-fit to the measured data points assuming a power law: $\omega(\theta) = A_\omega \theta^{-\delta}$ when the slope is set free. The dashed line is the best-fit when the slope is fixed at $\delta = 1$

where DD refers to the number of different cluster-cluster pairs, DR is the number of cluster-random pairs, and RR is the number of random-random pairs with a separation between $\theta - \delta\theta/2$ and $\theta + \delta\theta/2$. For a robust computation, we used $\sim$1000 times more random points ($N_R = 300,000$) than the number of candidate clusters ($N_D = 308$) over the same Boötes field. Figure 17 shows the measured ACF (Black circles). The error bars on the observed ACF are computed using 100 bootstrap resamplings of the candidate cluster sample with replacement. Assuming the observed angular correlation function follows a power law

$$\omega(\theta) = A_\omega \theta^{-\delta}, \tag{1.2}$$

we fit the function in equation (1.2) to the observed ACF by varying A_ω and δ simultaneously, with θ in units of degrees. Here A_ω and δ are the amplitude and slope of the power law, respectively. These parameters are covariant in nature. The covariance ellipses in Figure 8 shows the 1σ (light-orange), 2σ (orange), 3σ (red) errors, and covariances in A_ω and δ. Thus, we obtained the best-fit $logA_\omega = -1.51^{+0.37}_{-0.53}$ and $\delta = 1.11^{+0.46}_{-0.42}$. Considering the 1σ error on the best-fit slope, δ is consistent with previous cluster-autocorrelation studies [4, 14, 40, 67, 95], which generally obtain a slope close to unity. The best-fit model is represented by the dashed line in Figure 17. Due to clustering and finite survey area, the integral of the density contrast over a finite volume is not zero. This issue was corrected by adding an integral constraint (IC) to the observed ACF in equation (1.1). The Boötes field is wide enough to be minimally impacted by this correction factor.

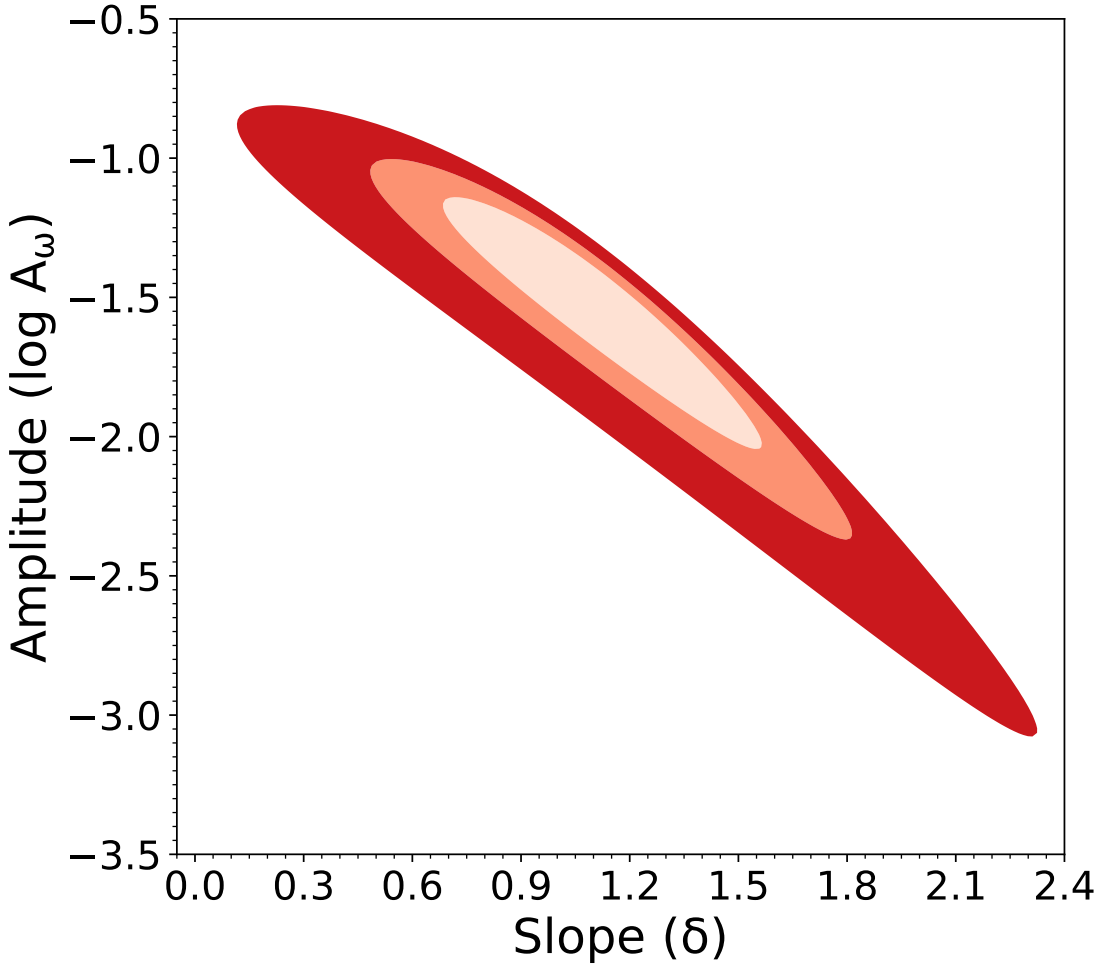

Figure 8: Ellipses showing the 1σ (68.2%, light-orange), 2σ (95.4%, orange), 3σ (99.7%, red) confidence regions, and covariances in A_ω and δ for the power-law fit of the measured angular correlation function.

However, following the numerical solution in Roche & Eales [106]

$$IC = A_\omega \frac{\sum RR(\theta)\theta^{-\delta}}{\sum RR(\theta)}, \tag{1.3}$$

we obtained IC ≈ 0.08 using the best-fit A_ω and δ. We obtained IC ≈ 0.09 by varying A_ω only, and holding $\delta = 1$ fixed. In both cases, the obtained values are negligible.

1.6 Discussion

1.6.1 Spatial Clustering Length

Spatial clustering length (r_o)—a distance at which the real-space correlation function (ξ_r) reaches unity—is a widely used indicator of the correlation strength. We computed r_o by deprojecting the ACF (Equation 1.2) over the redshift range [76] spanned by the DOGPaCS cluster candidates

$$r_o^\gamma = A_w \left\{ \frac{H_o H_\gamma}{c} \frac{\int_{z1}^{z2} N^2(z)[x(z)]^{1-\gamma} E(z) dz}{[\int_{z1}^{z2} N(z) dz]^2} \right\}^{-1} \tag{1.4}$$

Here $\gamma = 1+\delta$, $H_\gamma = \Gamma(1/2)\Gamma\{(\gamma-1)/2\}/\Gamma(\gamma/2)$, $E(z) = H(z)/H_o$, and $x(z)$ represents the comoving radial distances to the cluster candidates. The primary source of error in r_o is the uncertainty of the redshift distribution of the cluster sample $N(z)$. We computed $N(z)$ using the mean photometric redshifts of the cluster members and obtained the range of the DOGPaCS cluster redshifts: $1.3 \lesssim z \lesssim 1.8$. The inset plot in Figure 9 shows the redshift distribution of the cluster candidates. Finally, we obtained $r_o = 16.13\, h^{-1}$Mpc in comoving scale using the Limber inversion

formula in equation (1.4). A fiducial Gaussian with $\bar{z} = 1.55$ and $\sigma_z = 0.08$ also yields identical results. Strictly speaking, the true shape of $N(z)$ is unknown due to large photo-z errors of the SDWFS galaxies at $z > 1.3$ (Section 1.4.2). A narrower (broader) redshift distribution would reduce (increase) the clustering length. We derived $r_o = 16.13^{+8.04}_{-7.14}$ h^{-1}Mpc by varying A_ω at the best-fit slope ($\delta = 1.11$). At fixed $\delta = 1$, we obtained $r_o = 16.22^{+1.76}_{-1.61}$ h^{-1}Mpc. In both cases, the results are consistent with each other.

We computed the comoving volume density of the DOGPaCS cluster candidates $n_c = 1.70 \times 10^{-5}$ h^3Mpc^{-3}. Figure 9 illustrates the relationship between r_o and n_c for different cluster and protocluster surveys over a broad redshift range: $0 \lesssim z \lesssim 4$. Our result is consistent with the analytical prediction for the ΛCDM model [127], shown as the dashed line in Figure 9. Assuming the n_c-r_o relationship is invariant with redshift, this work resembles the clustering pattern of the ISCS clusters [14], SWIRE clusters [95] at similar redshifts. At low redshifts ($z \sim 0$), optically-selected, rich SDSS clusters [4, 40] show comparable results. This similarity suggests that the high-redshift DOGPaCS candidates are the progenitors of the present-day massive clusters.

1.6.2 Comparison with the ISCS Clusters $z > 1.3$

The ISCS is a pioneering cluster survey (see § 1.2.2) in Boötes. Galaxy clusters in the ISCS sample at $z \gtrsim 1$ (§ 1.2.2) are primarily overdensities of infrared-bright,

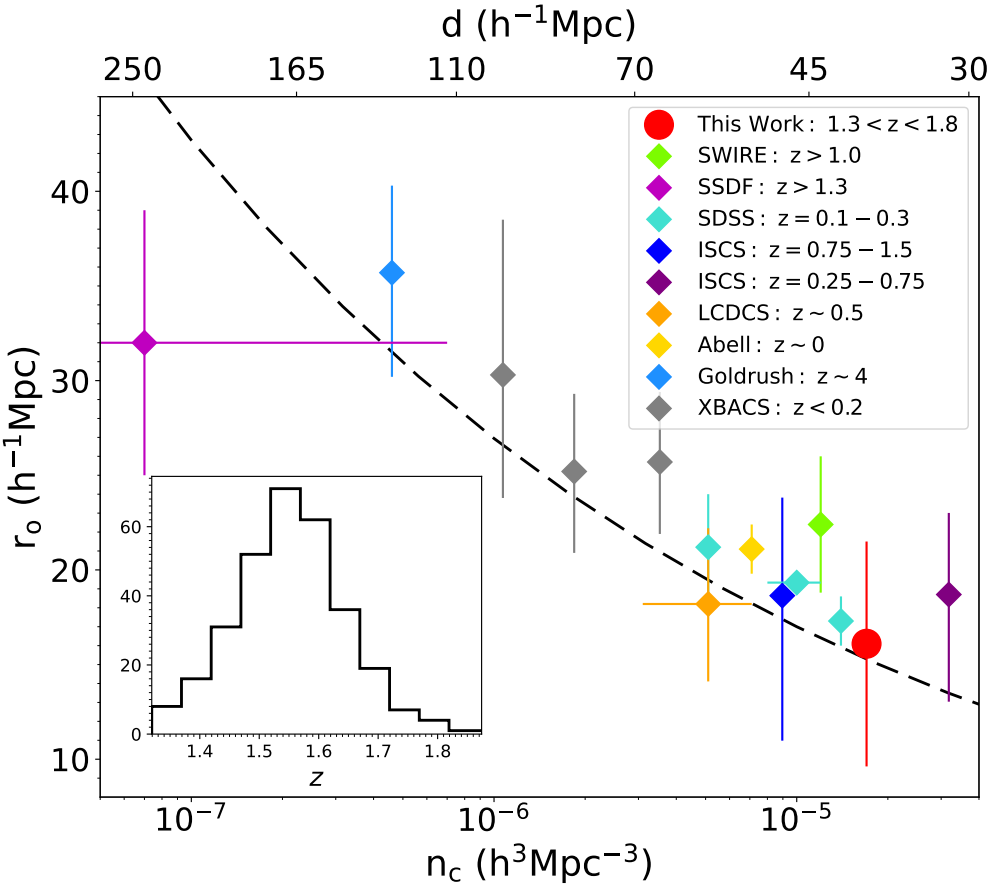

Figure 9: Comoving volume density (n_c) vs spatial clustering length (r_o) for various cluster and protocluster surveys. The cyan filled diamonds show the optical-selected SDSS clusters [4, 40]. The Abell cluster [98], X-ray clusters [74], and LCDCS cluster [50] are shown in yellow, grey, and orange diamonds. The purple and blue diamonds represent the ISCS intermediate- and high-redshift clusters [14], respectively. The SWIRE [95] and SSDF [104] high-z clusters are shown in green and magenta diamonds. The only protocluster sample at $z \sim 4$ is shown in light-blue diamond [116]. The red data point shows our work. An analytic approximation of the form, $r_o = 1.7(n_c)^{-0.2}$ for the ΛCDM n_c-r_o is shown using the dashed black line [127]. **Inset:** Histogram showing the photometric redshift distribution of the DOGPaCS cluster candidates at $1.3 \lesssim z \lesssim 1.8$.

i.e., massive galaxies in photometric redshift slices. The emergence of the red-sequence of passively evolving galaxies in many spectroscopically confirmed ISCS clusters at $z \gtrsim 1$ supports the previous statement [39]. In contrast, this work has identified a sample of cluster candidates selected exclusively by DOGs, i.e., star-forming galaxies or AGNs. The systematic biases are different in the two cluster surveys, yet it is intriguing to compare them as both cluster-searches have detections at $z > 1.3$. An ISCS cluster within 2 arcmin ($\sim$ 1 Mpc) of the DOGPaCS candidate with a difference smaller than 0.10(1+z) along the z-direction was considered a match, i.e., re-discovery.

Despite the differences in the two surveys, our method re-discovered 29 ISCS clusters, nearly 60% of all the ISCS clusters at $z \geq 1.3$ (Figure 6). There are a total of ten spectroscopically confirmed ISCS clusters at $z > 1.3$ [19, 39, 110, 129]. We re-discovered five of them with more than 4σ significance (Figure 10). Indeed, the top DOGPaCS cluster candidate, a $\sim 12\sigma$ significant detection, is the re-discovery of a spectroscopically-confirmed [18, 110], massive cluster ISCS J1438.1+3414 ($M_{200} \sim 2.24 \times 10^{14}\ M_{\odot}$) at $z_{spec} = 1.41$. The mean angular separation and redshift difference (Δz) between the ISCS and the DOGPaCS matches at $z \geq 1.3$ are $\sim 30 (\approx 250$ kpc), ~ 0.06, respectively. The corresponding values between the spectroscopically confirmed ISCS clusters and DOGPaCS matches are only $\sim 15 (\approx 130$ kpc) and ~ 0.03. Here Δz is the difference between the mean photo-z of the member galaxies of a DOGPaCS cluster and the spectroscopic redshift of the matched ISCS cluster, scaled by $1/(1+z)$. The entire matching results are listed in Table 1.4.4. It is not

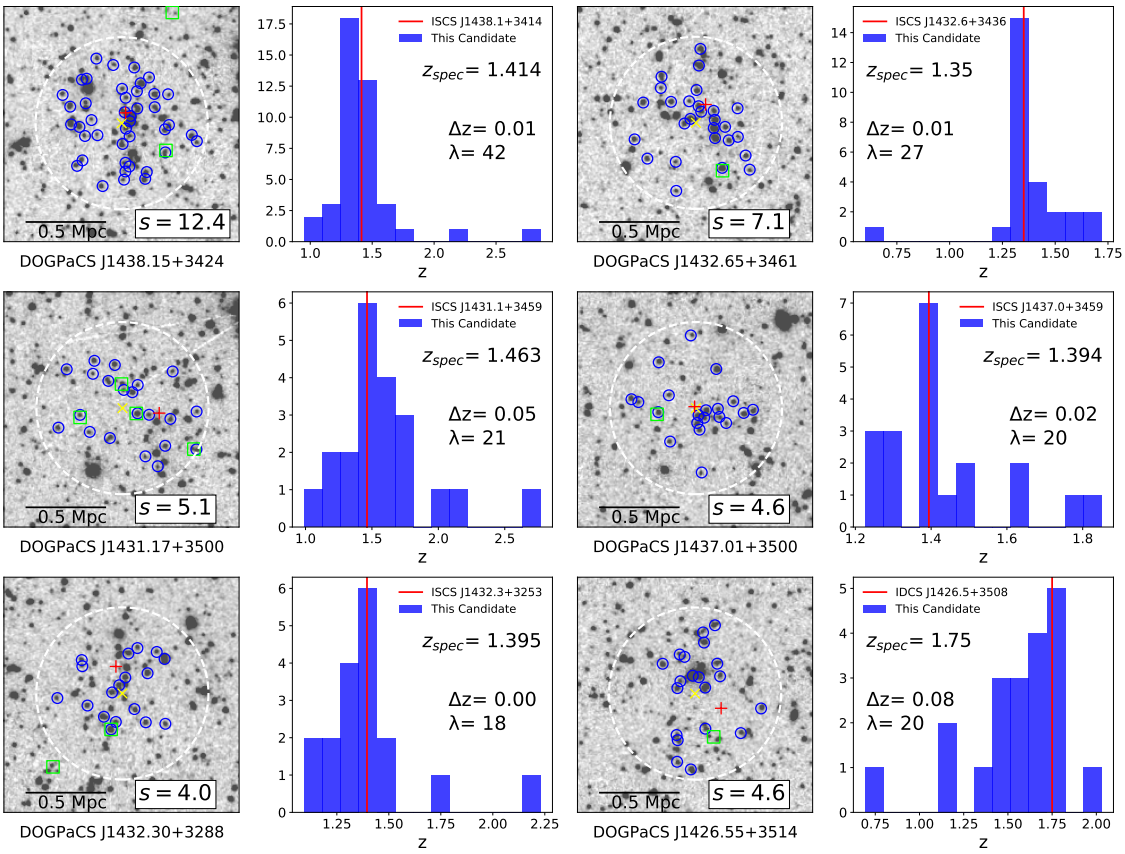

Figure 10: Illustration of the re-discovery of six (6) spectroscopically-confirmed ISCS and IDCS galaxy clusters at $z_{spec} \gtrsim 1.4$ with very high detection significance (s) and high richness value (λ). The bottom-right panel illustrates the re-discovery of the most massive galaxy cluster IDCS J1426.5+3508 at $z_{spec} = 1.75$ with $\sim 5\sigma$ detection significance. The IRAC 4.5μm maps illustrate the matching in the projected sky, while the adjacent histograms show the matching along the z direction. The dashed white circles represent the 1 arcmin search-radius used by our method. The green squares mark the locations of the DOGs (signposts). The member galaxies in each cluster, selected by our pipeline, are shown in blue circles. Our pipeline identified the centroids (yellow *cross*) of these clusters in close proximity ($\lesssim$ 150 kpc) to the centers (red *plus*) reported by the ISCS/IDCS survey. These clusters are matches in the projected sky and the redshift direction. The associated histogram with each 4.5 μm map shows the redshift distribution of the member galaxies. Although the member galaxies are selected by color and magnitude, the spectroscopic redshift of the cluster (solid red line) coincides with the peak of the redshift distribution of the members in all cases. The Δz value represents the difference between the cluster spec-z and the mean of the member galaxies' redshifts scaled by $1/(1+z)$.

surprising that our pipeline has not detected many other ISCS clusters. Most of these clusters either do not host any DOGs, or the detection is insignificant.

1.6.3 Re-discovery of IDCS 1426.5+3508: The Most Massive Galaxy Cluster at $z > 1.5$

In an endeavor to find distant galaxy clusters, the ISCS survey was extended using the SDWFS data products. This photo-z-based search campaign is known as the IRAC distant cluster survey (IDCS). Subsequently, two distant large-scale structures were identified: a mature cluster, IDCS J1426.5+3508 at $z = 1.75$ [111] and another (forming)cluster IDCS J1433.2+3306 at $z = 1.89$ [128]. The full search result will be published elsewhere. The IDCS J1426.5+3508 is the most massive (X-ray and SZ-detected), spectroscopically confirmed galaxy cluster at $z > 1.5$ to date. Our pipeline re-discovered the IDCS J1426.5+3508 cluster with a $\sim 5\sigma$ significance. The bottom-right panel in Figure 10 illustrates the re-discovery. We identified 20 IRAC high-z galaxies (blue circles) as members of this cluster. Our pipeline also identified the 4.5 μm flux-weighted centroid of the member galaxies (yellow cross), $\sim$ 300 kpc away from the tracer DOG in the north-east direction. This centroid is offset by only 15 ($\sim$ 125 kpc) from the X-ray centroid (red plus symbol) of the IDCS cluster [15]. Although we selected the high-z galaxies based on IRAC color and magnitude, the majority of the member galaxies have photometric redshifts consistent with the spectroscopic redshift of the IDCS cluster at $z = 1.75$ (redshift histogram, bottom-right panel, Figure 10). Again, this demonstrates the success of our method in identifying distant clusters at $z > 1.5$.

1.6.4 Mass and Evolution of the Cluster Candidates

We used the HMFcalc [88] and its successor HALOMOD [87], an online tool, to perform an abundance matching analysis. We generated a halo mass function (HMF) using the analytical prescription in Behroozi et al. [8]. HMF is the prediction of the number of dark matter halos per unit mass per unit comoving volume. The total number of halos occupying the comoving volume spanned by $z = 1.3 - 1.8$ was derived by computing the volume integral of the HMF over a range of halo mass ($10^{12} M_{\odot} h^{-1} \leq M \leq 10^{15} M_{\odot} h^{-1}$). Assuming a dark matter halo contains only one galaxy cluster, we matched the predicted number of halos by HMFcalc to the observed number of DOGPaCS cluster candidates (308). The area under the curve (shaded region) in Figure 11 represents the matched number of halos. This matching allowed us to compute the minimum mass of the halos that contain the DOGPaCS sample $M_{200,min} = 2.34 \times 10^{13} M_{\odot} h^{-1}$, where the halo overdensity is 200 times the mean density of the universe. We also used other HMFs [114] and obtained nearly identical minimum halo mass. This implies that the DOGs are sensitive tracers of overdensities down to forming cluster or protocluster core halos, still embedded in their birth environment. This is not surprising because early-stage clusters with small velocity dispersions produce higher galaxy-galaxy merging frequencies, i.e., DOG-like events [19]. Indeed, the merger rate is directly observed to be enhanced manifold in cluster galaxies than the field expectation at $z \geq 1.5$ [77, 120]. We used HALOMOD to compute the minimum halo mass of a fiducial cluster sample at $z = 0$ with a similar comoving volume density of the DOGPaCS

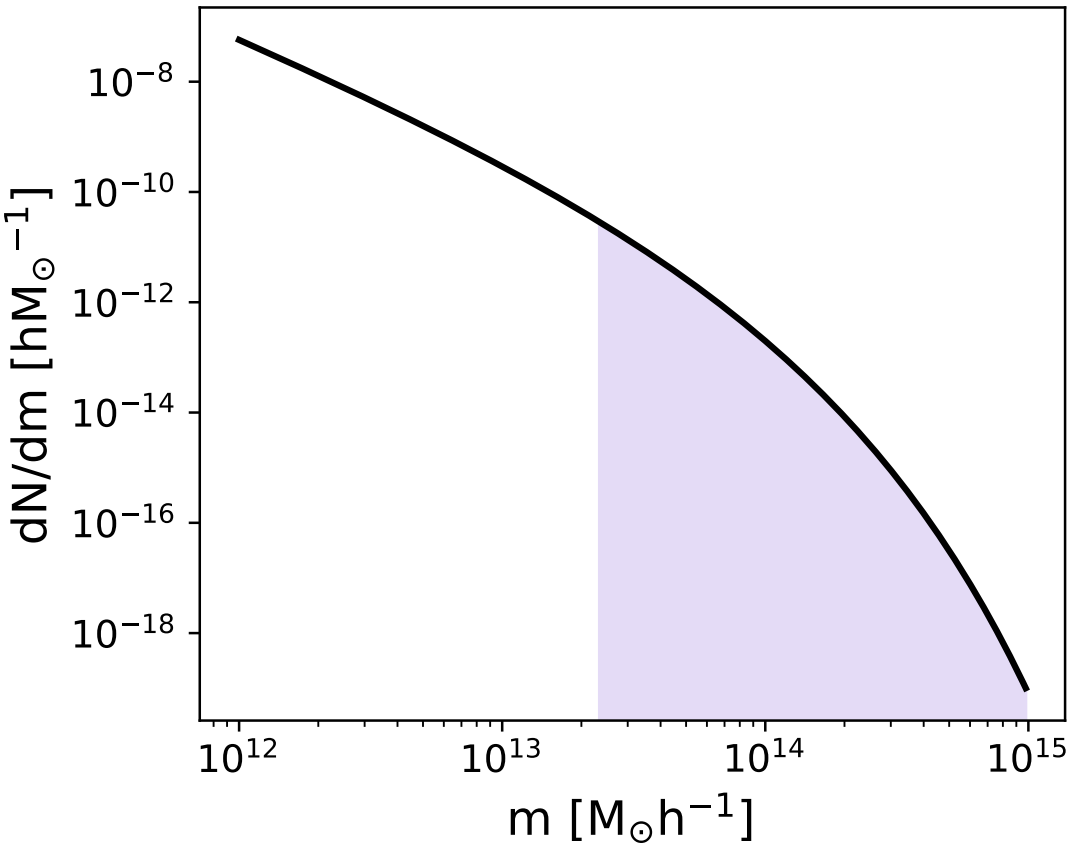

Figure 11: The HMFcalc prediction of the number of dark matter halos per unit mass at $z = 1.3 - 1.8$ using the prescription in Behroozi et al. [8], subtended by the solid angle in Boötes. The integral under the shaded region (purple) matches the observed number of DOGPaCS candidates.

sample, $M_{200,min}$= $1.38 \times 10^{14} M_{\odot} h^{-1}$. This suggests that the DOGPaCS clusters will evolve into massive clusters by $z = 0$, assuming all of them will grow individually without merging.

1.7 Summary

We have studied the relationship between a sample of Dust-Obscured Galaxies (DOGs) and a subset of the galaxy clusters at $z > 1.3$ from the IRAC Shallow Cluster Survey (ISCS) in Boötes, which has revealed a high concentration of the DOGs within $\sim$ 250 kpc around the ISCS clusters. Inspired by the success of the preliminary check, we employed the DOGs as signposts to find new galaxy clusters in the ~ 10 deg^2 Spitzer Deep, Wide-Field Survey (SDWFS) in Boötes. IRAC color and I-band magnitude were used to isolate the SDWFS galaxies at $z \gtrsim 1.3$, which we referred to as the IRAC high-z sample. We developed a cluster-detection algorithm using the DOGs as signposts, and ultimately, we identified 308 overdensities of the IRAC high-z galaxies, i.e., cluster candidates at $1.3 \lesssim z \lesssim 1.8$ with a significance greater than 3σ. Our pipeline re-discovered nearly 60% of the $z > 1.3$ ISCS clusters with a matching distance $\lesssim$ 250 kpc, including IDCS J1426.5+3508, the most massive cluster above $z > 1.5$. The X−ray centroid of IDCS J1426.5+3508 is only $\sim$ 125 kpc away from the centroid location identified by our pipeline.

We computed the two-point angular correlation function of the cluster candidates. The observed angular clustering follows a power law of the form: $\omega(\theta) = A_{\omega}\theta^{-\delta}$. The best-fit parameter values are as follows: $logA_{\omega} = -1.51^{+0.37}_{-0.53}$ and $\delta =$

$1.11^{+0.46}_{-0.42}$. The cluster candidates show a strong clustering length, $r_o = 16.13^{+8.04}_{-7.14}$ h^{-1}Mpc with a comoving volume density, $n_c = 1.70 \pm 0.01 \times 10^{-5}$ $h^3 Mpc^{-3}$. Using the abundance matching technique, we obtained the minimum mass of the dark matter halos that host the cluster sample, $M_{200,min} = 2.34 \times 10^{13} M_{\odot} h^{-1}$. Eventually, the cluster candidates will grow over time and evolve into structures with halo mass $\gtrsim 10^{14}$ $M_{\odot}$ halos by $z = 0$. This suggests that the DOGs are sensitive signposts of the progenitor population of the present-day massive clusters.

CHAPTER 2

A PROTO-SUPERCLUSTER CANDIDATE HOSTING A MASSIVE GALAXY CLUSTER AT z=1.75

Unpublished: currently under review

Abstract

We present a proto-supercluster candidate hosting the most massive, spectroscopically-confirmed galaxy cluster IDCS J1426.5+3508 at z=1.75 in Boötes. Using a sample of mid-infrared-bright Dust-Obscured Galaxies (DOGs) as signposts, we have identified 31 (10) galaxy cluster-scale overdense regions of photo-z selected galaxies that are significant at $\geq 3\sigma$ (4σ) level at z=1.75. In addition, we have also identified three protocluster-scale nodes with greater than 3σ detection significance in the filamentary structure. This suggests that the enormous structure is still undergoing the formation of new clusters. One of the cluster-scale nodes is the rediscovery of the IDCS cluster with more than 4σ significance. Together with the IDCS cluster, the cluster- and protocluster-scale nodes span an area of 50 Mpc $\times$ 35 Mpc on the physical scale. We have also identified approximately 23 spectroscopically confirmed AGNs at $1.72 \leq z_{spec} \leq 1.78$ from the AGES survey across the proto-supercluster. This is an outstanding example of a massive, X-ray/SZ-detected high-z galaxy cluster (IDCS J1426.5+3508) still embedded in its birth environment. We present the initial map of this remarkable multi-nodal structure and establish the opportunities

for future space- and ground-based multi-wavelength studies of the galaxy populations and ICM dynamics across the large-scale structure.

2.1 Introduction

Recent deep and wide photometric surveys have detected primordial, filamentary supercluster structures in which several protoclusters are embedded [29, 54, 115] at $z = 2.5 - 7.0$. These extreme examples may represent the progenitors of the supercluster structures found in the low and intermediate redshifts, such as the Lynx supercluster at $z = 1.3$ [107], CL1604 at $z = 0.9$ [55], the eFEDS supercluster at $z = 0.36$ [47], and the local Laniakea supercluster at $z = 0$ [118]. Although it is theoretically possible to associate a high-z proto-structure to its descendent $z = 0$ virialized halo, the large uncertainties in their accretion history prevent drawing a meaningful evolutionary connection between primordial superclusters (protoclusters) and superclusters (clusters) [23]. Simulation studies show that for a given cluster mass at $z = 0$, the mass of its main progenitor halo can vary widely [86]. This stochasticity in the evolutionary states can be explained only by identifying and characterizing large samples of protoclusters at $z \sim 2$ during their formation epoch of the present-day massive clusters.

Protocluster discoveries were offshoots of narrow-band spectroscopic surveys before the advent of wide-field imaging facilities [66, 112]. Distant wide-field optical imaging has produced large samples of candidate protoclusters at $z > 3$ using

the dropout technique as the redshifted Lyman-limit falls between two imaging filters [116]. Spectral absorption line studies of the background bright quasars in the line-of-sight of protoclusters are also useful in finding overdense structures [28, 48]. Protoclusters are also often found associated with radio galaxies and QSOs [92, 125]. However, the emission line or radio-selected protocluster searches often miss out on the crucial step of galaxy evolution, i.e., the obscured star-formation phase in the protocluster galaxies. Planck Collaboration et al. [102] have identified overdensities of unresolved, far infrared-bright cold sources, i.e., Dusty-Star Forming Galaxies (DSFGs). Similar overdensities of dusty starbursts have also been found by other studies [21, 31]. These discoveries are not surprising because galaxy clusters acquire the bulk of their stellar mass via dusty, obscured star-formation in the protocluster phase [25]. Systematic searches using accurate photometric redshifts have also detected large samples of protoclusters at $z > 2$ [24, 75].

In this study, we report the discovery of a proto-supercluster candidate at $z = 1.75$ in the field of Boötes. This multi-component filamentary structure is a pioneering example in the distant universe that hosts both clusters and protoclusters, including IDCS J1426.5+3508, the most massive galaxy cluster at $z = 1.75$. The clusters and the protocluster candidates are overdense regions of photometric redshift-selected galaxies at $z = 1.75$, identified using a sample of optically-faint, Ultra-Luminous Infrared Galaxies (ULIRGs) called the Dust-Obscured Galaxies (DOGs) as signposts. DOGs are exclusively high-z dusty star-forming galaxies

(SFGs) and/or AGNs at $z \sim 2 \pm 0.5$ with no low-z counterparts [34]. Leveraging the fact that the dusty SFGs and AGNs are abundant in galaxy clusters and protoclusters at $z > 1.5$ [21, 92], the Dust-Obscured Galaxy Cluster and Protocluster Survey (DOGPaCS) has demonstrated the efficacy of using DOGs as signposts to identify clusters at $1.3 < z < 1.8$ in a companion paper (Saha et al. 2022a in preparation). As part of the DOGPaCS campaign, this study has identified an enormous proto-supercluster candidate at $z = 1.75$ by exploiting a multi-wavelength data set in Boötes.

We present the data used in this study in §2.2. We describe the (proto) cluster-finding algorithm and report the discovery of a proto-supercluster candidate §2.3. We discuss and characterize our detection in §2.4. We use the ΛCDM benchmark model throughout the paper. The values of the following cosmological parameters are used: $\Omega_m = 0.3$, $\Omega_\Lambda = 0.70$, and $H_0 = 70$ kms^{-1}Mpc^{-1}.

2.2 Data

2.2.1 A Multi-wavelength Catalog in Boötes

Over the past two decades, the ~ 9 deg^2 field of Boötes has been observed in multiple wavelengths ranging from X-ray to radio wavelengths. For the photometric redshift (photo-z) estimation purpose, we used a 12-band multi-wavelength catalog from Duncan et al. [37] (hereafter D18 catalog), spanning a wavelength range from 0.36 to 8.0 μm. This catalog includes optical photometry in B_w, R, and I bands from the NOAO Deep Wide-Field Survey [NDWFS, 20, 65]. The photometry in

the U- and Y-band were provided by the Large Binocular Telescope (LBT) Boötes field survey [10]. NEWFIRM imaging of the Boötes field was explored to extract detections in near-infrared J, H, K_s bands [51]. Finally, the IRAC [44] mid-infrared observations in 3.6, 4.5, 5.8, 8.0 μm were incorporated from the *Spitzer* Deep Wide-Field Survey [3]. The full matching procedure and detailed aperture photometry are provided in Duncan et al. [37]. We selected only sources with *signal-to-noise* ratio (SNR) ≥ 5 in 4.5 μm to reduce noisy selection and obtain better photo-z estimates $z > 1.5$. The final D18 catalog contains $\approx 4 \times 10^5$ sources in the Boötes field.

2.2.2 Selection of Galaxies at $z = 1.75$

2.2.2.1 Photometric Redshifts

We computed the photo-z of the sources in the final D18 catalog using the 12-band photometry. We adopted a template-fitting photo-z code, described in Brodwin et al. [16], and computed accurate redshift probability distribution function, P(z), using a set of empirical templates [103]. The addition of the near-infrared photometry has produced better quality photometric redshifts for the sources in the wide Boötes field. To measure the accuracy of the photometric redshifts, we matched the D18 catalog to the best available spectroscopic redshifts in Boötes obtained by several studies over the years, such as the AGN and Galaxy Evolution Survey [AGES, 69] and many cluster studies [19, 39, 111, 129]. The matched

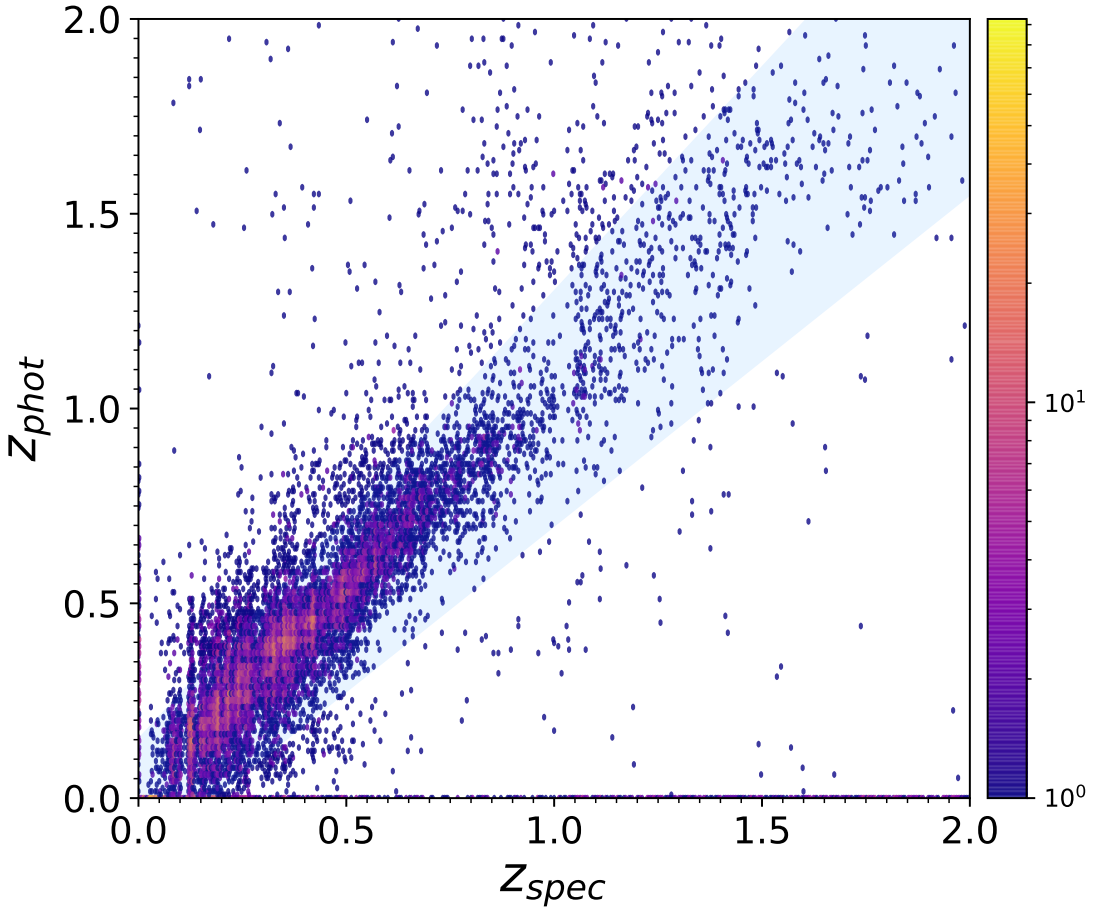

Figure 12: Hexagonal binning plot of z_{phot} versus z_{spec} for the D18 sources with spectroscopic redshifts. The color bar on the right indicates the number of sources in the bins. The region, enclosed by the $z_{phot} = z_{spec} \pm 0.15 \times (1 + z_{spec})$ lines, is shown in light blue. Most of the sources at $z > 1.5$ are mainly spectroscopically confirmed AGNs from the AGES survey [69].

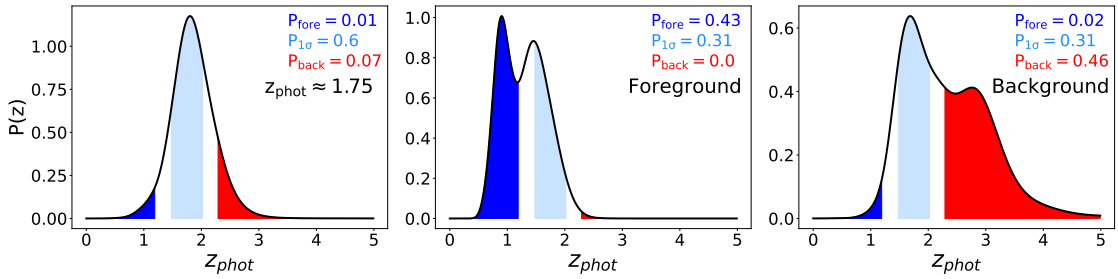

Figure 13: The area under the P(z) curve (solid line) within $\pm 1\sigma$ of $z = 1.75$ ($P_{1\sigma}$) is shown in light-blue. The red and blue represent the areas under the P(z) curve outside of the -2σ and +2σ boundaries about $z = 1.75$, respectively. **Left:** A galaxy at $z_{phot} = 1.75$ that satisfies Equations (2.1), (2.2), and (2.3). **Middle:** A possible foreground object with $P_{fore} > 0.3$ which satisfies equation (2.1) only. **Right:** A possible background object with $P_{back} > 0.3$ that meets the condition in Equation (2.1) only.

catalog contains $\approx$ 20500 D18 sources with spectroscopic redshifts within a matching offset smaller than 1. This catalog was used for calibrating the photo-z results. Figure 12 shows a comparison between the peak of the redshift probability distribution (z_{phot}) and the true redshift (z_{spec}). Quantitatively, we obtained a photometric redshift error, $\sigma/(1+z) \approx 0.06$ at $z \lesssim 1$. The corresponding value is ≈ 0.10 for sources at $z = 1.5 - 2.0$, where most of the objects in the calibration catalog are primarily AGNs from the AGES survey. Here, σ is the normalized median absolute deviation (NMAD) of the residuals, defined as $\sigma_{\rm NMAD} = 1.4826 \times median(|\Delta z - median(\Delta z)|)$ [41]. The residual (Δz) is the difference between z_{phot} and z_{spec}, scaled by $1/(1+z_{spec})$. $\sigma_{\rm NMAD}$ is a better estimator of the photometric redshift error than the rms dispersion, as it does not take catastrophic outliers into account.

2.2.2.2 The $z175$ sample

We exploited the full normalized probability distribution function, P(z), of the photometric redshift for a source to detect the objects at $z = 1.75$. We imposed three criteria to isolate objects that have photometric redshifts consistent with $z = 1.75$:

$$\mathrm{P}_{1\sigma} = \int_{z_s-\sigma}^{z_s+\sigma} \mathrm{P(z)}\ \mathrm{d}z \geq 0.3 \tag{2.1}$$

$$\mathrm{P}_{\mathrm{fore}} = \int_{0}^{z_s-2\sigma} \mathrm{P(z)}\ \mathrm{d}z < 0.3 \tag{2.2}$$

$$\mathrm{P}_{\mathrm{back}} = \int_{z_s+2\sigma}^{\infty} \mathrm{P(z)}\ \mathrm{d}z < 0.3 \tag{2.3}$$

Here $z_s = 1.75$, and $\sigma = 0.10(1 + z_s)$. We isolated $\approx$ 47700 objects that satisfied the condition in Equation (2.1) in Boötes. For our purposes, objects for which the area under the P(z) curve within $\pm 1\sigma$ of $z = 1.75$ is 30% or more ($\mathrm{P}_{1\sigma} \geq 0.3$) are considered objects with photometric redshift, $z_{phot} = 1.75$ [32]. This selection provides a balance between obtaining a pure versus complete sample at $z = 1.75$. The left panel in Figure 13 shows an example of a galaxy at $z_{phot} = 1.75$ with $\mathrm{P}_{1\sigma} \geq 0.3$ which has a very little or no probability to be outside of $z_s \pm 2\sigma$. However, the condition alone in Equation (2.1) also yielded false positives as shown in the middle and right panel in Figure 13. Therefore, we also applied conditions in Equation (2.2) and (2.3) to remove the *line-of-sight* foreground and background interlopers from the sample, respectively. The final sample contains $\approx$ 39300 objects at $z_{phot} = 1.75$ in Boötes. This sample will be referred to as the $z175$ sample.

2.2.3 Dust-Obscured Galaxies (DOGs)

Using the *Spitzer* 24 μm observation of the Boötes field, Dey et al. [34] identified a sample of $\approx$ 2600 extremely dusty and infrared bright galaxies called Dust-obscured Galaxies (DOG) at $z \sim 2$. DOGs are transitory stages between submillimeter and passively evolving galaxies that are undergoing either or both extreme star formation and AGN accretion. Redshifts are unknown for most DOGs. However, a subset of these DOGs ($\sim$ 100) has been confirmed spectroscopically at $z \approx 2.0 \pm 0.5$ [for details, see 34]. This study searched for overdensities of galaxies in the $z175$ sample by analyzing their distribution in the vicinity of the DOGs in the projected sky.

2.3 A Filamentary Proto-Supercluster Candidate at $z = 1.75$

2.3.1 Algorithm

We began by computing the number of $z175$ galaxies around each DOG in our sample via a *counts-in-cell* [45, 125] method. However, DOGs may not always reside in the central region of the high-z galaxy clusters. In a theoretical merger-driven scenario [19, 58], short-lived DOG-like phases are more likely to occur away from the center of the forming clusters, where the velocity dispersion is small and thus enhances the frequency of merging activities. Therefore, we designed and implemented a pipeline to identify the regions with the highest number of $z175$ galaxies (richness, λ) in a circular cell near the DOGs. Although the algorithm started by counting around the DOGs, it eventually identified local maximum densities of the

$z175$ galaxies in circular cells by moving the cells in the vicinity of the DOGs. The details of the search algorithm are described in a companion paper (Saha et al. 2022a in preparation). A cluster-scale radius, r = 1.5 (2.1 cMpc), was chosen as the cell-radius. The highest number of $z175$ galaxies in a circular cell is defined as the richness (λ) near a DOG. Given the redshift of our interest, many overdense regions are, perhaps, the main halos of forming clusters or protoclusters. Therefore, we conducted another search using a larger search radius of 3.2. This corresponds to 4.5 cMpc in comoving units. According to simulations [23], this scale roughly represents the size of protoclusters at $z = 1.75$ that will evolve into $\sim 3 - 10 \times 10^{14}$ $M_{\odot}$ or *Virgo-type* clusters by $z = 0$.

2.3.2 Overdensity Detection

For the cluster-scale search, the richness distribution near the DOGs is shown in Figure 14 (orange). This distribution was compared to the distribution of $z175$ galaxies around a sample of randomly distributed 200,000 points (dotted black line in Figure 14) to identify overabundances at $z = 1.75$. The horizontal offset between the two distributions implies that the DOGs, in general, prefer to reside in high-density regions compared to a randomly distributed population. We fit a Gaussian to the left-half of the random profile to obtain the true random expectation (μ) of the $z175$ galaxies in a circular cell. For the cluster-search, the best-fit Gaussian yielded $\mu \approx 8.5$ and dispersion, $\sigma \approx 3.1$. The corresponding values for the protocluster-scale search were 40.3 and 8.8, respectively. Thus, overdense regions with richness values

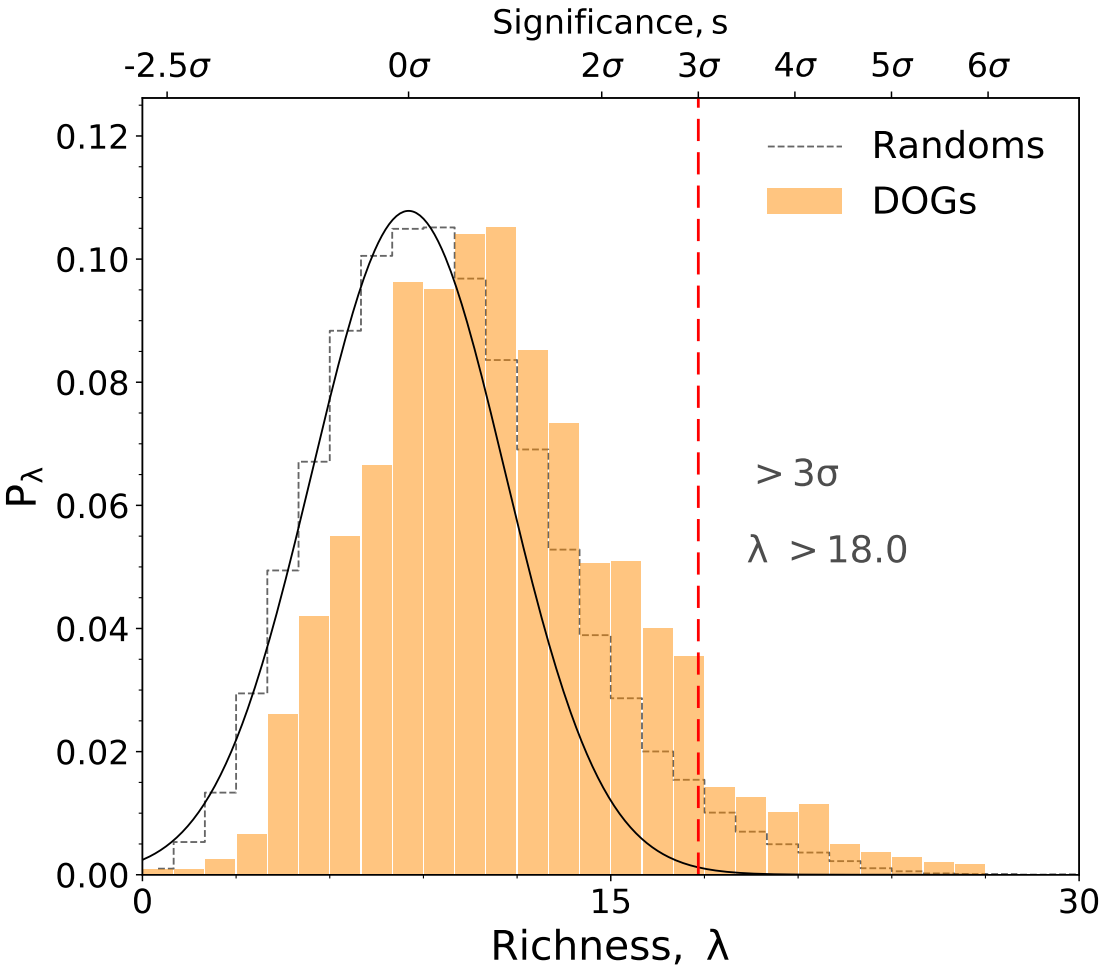

Figure 14: The orange histogram shows the richness — the highest number of galaxies in a circular cell of 1.5 ($\sim$ 2.1 cMpc) at $z_{phot} = 1.75$ — probability distribution near the DOGs in Boötes. The corresponding distribution for the randoms is shown in a dotted black line. The solid black line is the best-fit Gaussian to the random profile. Here, σ represents the standard deviation of the Gaussian. The top x-axis represents the richness values in the units of significance.

$\geq \mu + 3\sigma \approx 18$ (67) near DOGs were defined as the DOG-signposted cluster (protocluster) candidates with detection significance $\geq 3\sigma$ at $z = 1.75$. We considered cells with 80% or more good pixels for cluster or protocluster candidacy to reduce noisy detections. Good pixels were selected as pixels in a *Spitzer* mask image [3], with no imaging artifacts, such as diffraction spikes, saturation bleeding, or bad image columns.

2.3.3 The DOGPaCS Proto-Supercluster

The search for DOG-signposted (proto)clusters identified a remarkable large-scale structure comprising tens of galaxy clusters and three protocluster candidates at $z_{phot} = 1.75$, covering an area of 100×70 or 130×100 cMpc in comoving units in Boötes. The giant structure contains 31 DOG-signposted cluster candidates identified with greater than 3σ significance. Ten of these cluster candidates were identified with greater than 4σ significance, including IDCS J1426.5+3508. The IDCS cluster is the most massive galaxy cluster at $z > 1.5$ from the IRAC Distant Cluster Survey, with an X-ray mass of 2.3-3.3 $\times 10^{14}$ $M_{\odot}$ [15] and seven spectroscopically confirmed cluster members within 2 Mpc of the cluster center at $z_{spec} = 1.75$ [111]. In Figure 15, the 3σ (4σ) cluster candidates are shown in dashed white (yellow) circles. These substructures/nodes have formed the filaments of a web-like supercluster. The nodes are broadly stretched along three filamentary arms: two parallel north-south arms at R.A $\approx 217.0°$ and R.A $\approx 217.5°$, and a disjointed arm containing the IDCS cluster.

The search on a protocluster scale yielded three protocluster candidates with

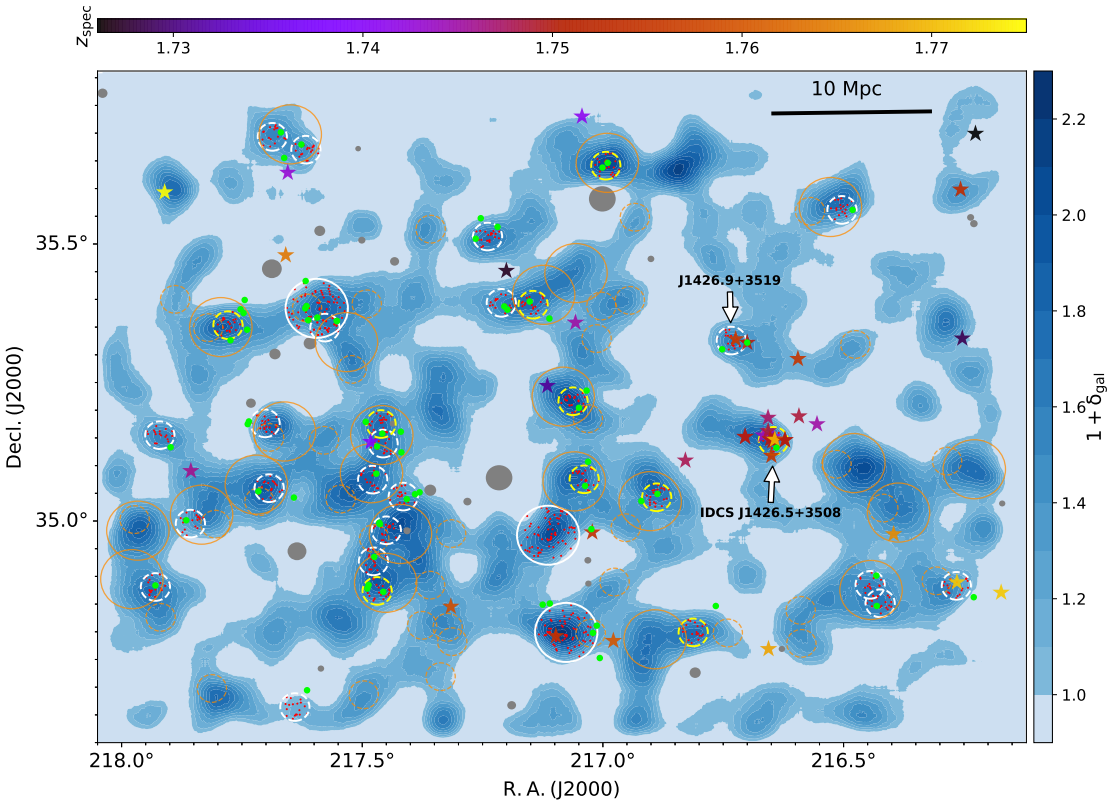

Figure 15: The blue contours represent the $\sim 50 \times 35$ Mpc smoothed density map of the photo-z selected galaxies at $z = 1.75$. The color bar on the right indicates the excess surface density of $z175$ galaxies (see §2.4). The darkest contour represents the most overdense regions, while the lightest blue color represents the underdense regions. Dashed yellow circles mark the most significant clusters with greater significance than 4σ. The dashed (solid) white circles show the distribution of the DOG-signposted cluster (protocluster) locations with significance $\geq 3\sigma$. The dashed/solid orange circles show the distribution of the DOG-signposted $\sim 2\sigma$ nodes for two different search scales. Small solid green circles indicate the location of the DOGs that signpost the associated nodes at $z = 1.75$. The small red dots represent the members of the nodes within the search radius of each centroid with photometric redshifts consistent with $z = 1.75$. The signposts for the $\sim 2\sigma$ nodes and member galaxies in them are not shown for clarity. All of these nodes map out the web-like filamentary structure. The gray circles mask the stars. The star symbols represent optically bright AGNs with known spectroscopic redshifts, where the colorbar indicates the redshifts at the top.

significance greater than 3σ, indicating that the supercluster is undergoing cluster formation at these nodes. Therefore, we refer to this structure as a forming supercluster or proto-supercluster. The candidate clusters and protoclusters in proto-supercluster are listed in Table 4. The locations of the protocluster are shown in solid white circles in Figure 4. One of the protocluster candidates located at the R.A $\approx 217.0^{\circ}$ filament is also significant on a cluster scale. This detection suggests that this cluster is still embedded in its pre-collapse environment. Overdense regions near the DOGs with significance smaller than 3σ were not deemed cluster or protocluster candidates. However, the $\sim 2\sigma$ overdensities, shown in orange circles, play an important role in understanding the true extent of the structure. Almost all $\geq 3\sigma$ nodes are associated with one or more relatively less dense DOG-signposted $\sim 2\sigma$ overdense regions (dashed orange circles in Figure 15) throughout the field. This suggests that the nodes are connected via the low-density regions and form a filamentary, web-like structure. Spectroscopic redshifts from the AGN and Galaxy Evolution Survey [AGES, 69] within $\Delta z = 0.03$ or $\Delta v \approx 3000$ kms^{-1} of the systemic redshift of the structure are shown as large stars where the colorbar at the top in Figure 15 indicates the redshifts. Spectroscopic observations near the IDCS cluster were obtained from Stanford et al. [111]. Most of the nodes are signposted by more than one DOG, as indicated by the solid green circles in Figure 15.

Table 4. List of the Nodes in the Proto-supercluster

Cluster/Protocluster	R.A. (J2000)	Decl. (J2000)	Richness (λ)	Significance (in σ)
DOGPaCS J1428.2+3512	217.056435	35.214089	24.0	5.0
DOGPaCS J1429.8+3510	217.458425	35.174845	23.0	4.7
DOGPaCS J1427.5+3502	216.883305	35.039516	23.0	4.7
DOGPaCS J1427.2+3447	216.809939	34.795042	22.0	4.4
DOGPaCS J1431.1+3521	217.778696	35.353775	22.0	4.4
DOGPaCS J1427.9+3538	216.983588	35.638763	22.0	4.4
DOGPaCS J1428.1+3504	217.034369	35.073056	21.0	4.0
DOGPaCS J1429.9+3452	217.467911	34.873496	21.0	4.0
DOGPaCS J1428.6+3523	217.138601	35.388807	21.0	4.0
DOGPaCS J1426.6+3508	216.638143	35.139263	21.0	4.0
DOGPaCS J1425.7+3450	216.419753	34.843716	20.0	3.7
DOGPaCS J1428.9+3530	217.233282	35.51236	20.0	3.7
DOGPaCS J1430.8+3503	217.692775	35.059393	20.0	3.7
DOGPaCS J1429.8+3458	217.448227	34.982934	20.0	3.7
DOGPaCS J1430.8+3510	217.701043	35.176458	20.0	3.7
DOGPaCS J1430.5+3540	217.61458	35.670254	20.0	3.7
DOGPaCS J1429.8+3508	217.454354	35.139489	20.0	3.7
DOGPaCS J1426.9+3519	216.723763	35.321092	19.0	3.4
DOGPaCS J1431.7+3509	217.921364	35.155126	19.0	3.4
DOGPaCS J1428.8+3523	217.205279	35.39255	19.0	3.4
DOGPaCS J1431.7+3452	217.930175	34.881075	19.0	3.4
DOGPaCS J1429.9+3504	217.476378	35.074479	19.0	3.4
DOGPaCS J1425.8+3452	216.440127	34.877291	19.0	3.4
DOGPaCS J1430.6+3439	217.63989	34.664236	18.0	3.1
DOGPaCS J1431.4+3459	217.857294	34.995529	18.0	3.1
DOGPaCS J1429.7+3502	217.412673	35.04369	18.0	3.1
DOGPaCS J1429.9+3455	217.475439	34.926911	18.0	3.1
DOGPaCS J1426.0+3533	216.488183	35.554501	18.0	3.1
DOGPaCS J1430.3+3520	217.575386	35.349105	18.0	3.1
DOGPaCS J1430.7+3541	217.68517	35.693483	18.0	3.1
DOGPaCS J1425.0+3452	216.260468	34.873287	18.0	3.1
DOGPaCS-PC J1430.4+3523	217.592479	35.383891	72.0	3.6
DOGPaCS-PC J1428.3+3447	217.074402	34.796872	71.0	3.5
DOGPaCS-PC J1428.4+3458	217.11049	34.972527	70.0	3.4

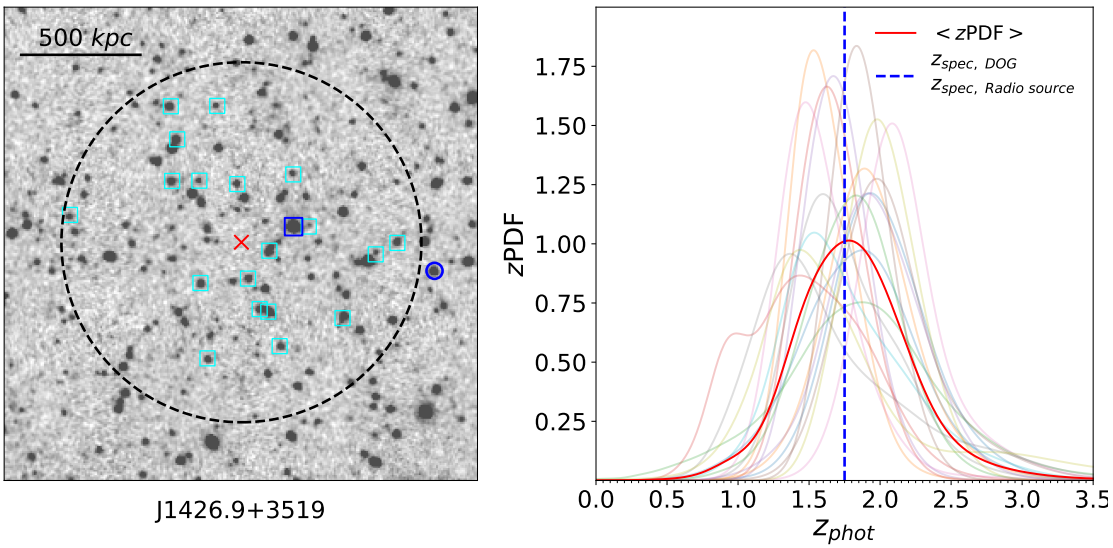

Figure 16: **Left**: The *Spitzer* IRAC 4.5 μm map of the cluster candidate DOGPaCS J1426.9+3519. The cyan squares show the locations of the photo-z-selected members of the cluster at $z = 1.75$, and the red cross is the 4.5 μm flux-weighted centroid of the member galaxies. The dashed black circle represents the 1.5 (2.1 cMpc) search radius. The blue circle denotes the location of the spectroscopically confirmed DOG at $z_{spec} = 1.75$, which signposts the cluster. The blue square near the center is a radio-bright AGN confirmed spectroscopically from the AGES survey at $z_{spec} = 1.75$. **Right**: The redshift probability distribution (zPDF) for the member galaxies is shown in various colors. The average zPDF is shown in a solid red line, and the spectroscopic redshift of the DOG is shown in a dashed blue line.

2.4 Discussion & Conclusions

The nodes in the proto-supercluster can be classified into three broad categories. The first kind is the majority of cluster-scale ($> 3\sigma$) nodes identified as protocluster-scale overdensities at the $\sim 2\sigma$ level as well. We speculate that these nodes are clusters-in-formation with infalling galaxies at the outskirts. The most prominent example of this kind is a $> 3\sigma$ protocluster candidate, DOGPaCS-PC J1430.4+3523 (see Figure 15) hosting a cluster-scale overdensity at R.A. $\approx 217.6°$. The second kind is the protocluster candidates that do not host such compact overdense regions, suggesting an early stage of cluster formation. We identified two protocluster nodes of this kind. One of these protocluster candidates (DOGPaCS-PC J1428.3 + 3447) hosts a bright AGN confirmed by spectroscopic analysis at $z = 1.75$, which is commonly found in high-z protocluster studies [71]. Finally, the third kind is the galaxy clusters that do not have an excess density on the protocluster scale, such as the IDCS 1426.5+3508. The IDCS cluster was rediscovered using the DOG-signpost method with 21 photo-z members from the $z175$ sample and 4σ detection significance. We also identified another similar cluster candidate, DOGPaCS J1426.9+3519, only ~ 8 Mpc away from the IDCS cluster. These clusters are examples of a handful of substructures in the proto-supercluster, implying a compact, mature stage of cluster evolution. Indeed, the IDCS cluster was observed to build a red sequence of galaxies, although with large uncertainties [111].

This study has identified a proto-supercluster with components in diverse stages of cluster evolution. However, the whole picture of this complex structure

might still be missing due to two main reasons. First, the uncertainty in the photo-z measurements at $z = 1.75$ yields broad redshift probability distribution functions (zPDF). Despite adopting the pruning techniques mentioned in §2.2.2.1, galaxies with overlapping PDFs from the neighboring redshifts appear at $z = 1.75$. A high level of contamination may artificially enhance the actual overdensity signal. Contrarily, galaxies at $z = 1.75$ that are not in $z175$ sample due to large photo-z errors can smooth out the density peaks. More accurate photo-z or follow-up spectroscopic observation of the individual nodes is required to confirm the three-dimensional shape of the proto-supercluster. Second, the tracer technique used in this study introduced a bias towards selecting the overdensities near the DOGs only. We constructed an excess density map of the $z175$ sample of galaxies across the proto-supercluster field to test this bias. The excess galaxy density, $\delta_{\rm gal} = ({\rm n_{gal}}/{\rm n_{avg}}) - 1$, for the $z175$ sample was computed, where $\rm n_{gal}$ and $\rm n_{avg}$ are the surface density of the photo-z-selected galaxies in the $z175$ sample and the average surface density of the $z175$ sample. The contours in Figure 15 show the excess density, where the lightest blue contours show the underdense regions in the proto-supercluster with a density smaller than the average galaxy density at $z = 1.75$. The DOG-signpost method identified most density peaks, i.e., the darkest blue contours, as cluster or protocluster nodes. Only a few of the peaks were missed by this study, mainly because these peaks do not host any DOGs. However, the basic filamentary structure emerged in our pipeline without using a galaxy density map.

2.4.1 DOGPaCS J1426.9+3519

This study identified the proto-supercluster nodes using the projected distribution of the DOGs. Strictly speaking, the association of the DOGs to the photo-z-selected overdensities at $z = 1.75$ is required to confirm via follow-up observations. Of the signposting DOGs, all but one do not have known redshifts. DOGPaCS J1426.9+3519 is of particular interest because this is the only cluster candidate ($\sim 3\sigma$) in the proto-supercluster where the signpost DOG is spectroscopically confirmed at $z = 1.75$. DOGPaCS J1426.9+3519 and IDCS J1426.5+3508 are neighboring clusters, located only ~ 8 Mpc apart. Figure 15 shows the locations of both clusters.

The *Spitzer* IRAC 4.5 μm map of DOGPaCS J1426.9+3519 is shown in Figure 16 (left panel). Our pipeline identified 19 photo-z-selected galaxies (cyan squares in Figure 16) as the cluster members within 1.5 of the cluster center at $z = 1.75$. The cluster centroid was detected only $\sim$ 100 away from the DOG ($\alpha = 216.690458^\circ, \delta = 35.317061^\circ$). The DOG signpost is a mid-infrared bright galaxy with *Spitzer* 24 μm flux density, f_{24}=1.142 mJy [34]. This cluster candidate also hosts an optically bright AGN confirmed spectroscopically from the AGES survey at $z = 1.75$ located near the cluster core. The Faint Images of the Radio Sky at Twenty Centimeters (FIRST) survey [7] using the NRAO Very Large Array (VLA) also reported this AGN ($\alpha = 216.714475^\circ, \delta = 35.323522^\circ$) as a bright radio source with peak flux density, $f = 4.58 \pm 0.15$ mJy. The right panel in Figure 16 shows the zPDFs of the member galaxies. The mode of the average zPDF coincides

with the spectroscopic redshift of the DOG demonstrating the success of the DOG-signpost method.

2.4.2 Clustering of DOGs in the Proto-Supercluster

Measurement of the correlation function is a common way to understand the clustering properties of constituent galaxies in a survey. We divided the Boötes field into two regions : the proto-supercluster region and the rest of the Boötes area outside the proto-supercluster. The two-point autocorrelation function (ACF) of the DOGs was computed in the two regions on angular scales, $\theta < 1^{\circ}$. The ACF measures the excess probability of finding a galaxy near another galaxy with respect to the random expectation. If a strong association between the DOGs and the proto-supercluster exists, it will be reflected in the clustering signal of the DOGs. Of the $\approx$ 2600 DOGs in our sample, nearly 450 *line-of-sight* DOGs reside in the projected sky of the proto-supercluster region. The ACF of all *line-of-sight* DOGs in the proto-supercluster was computed using a numerical estimator [53, 78]. The red data points (diamonds) in Figure 17 represent the ACF of the DOGs in the proto-supercluster region. The error in the measurements was computed using the bootstrapping method. We repeated the computation for the rest of the DOGs outside the proto-supercluster region in Boötes (black circles in Figure 17). A higher correlation signal was obtained for the DOGs in the proto-supercluster region, particularly at $\theta \lesssim 0.6$ ($\sim$1 cMpc). The stronger clustering pattern of DOGs indicates that DOGs reside in rarer and richer environments in the proto-supercluster region,

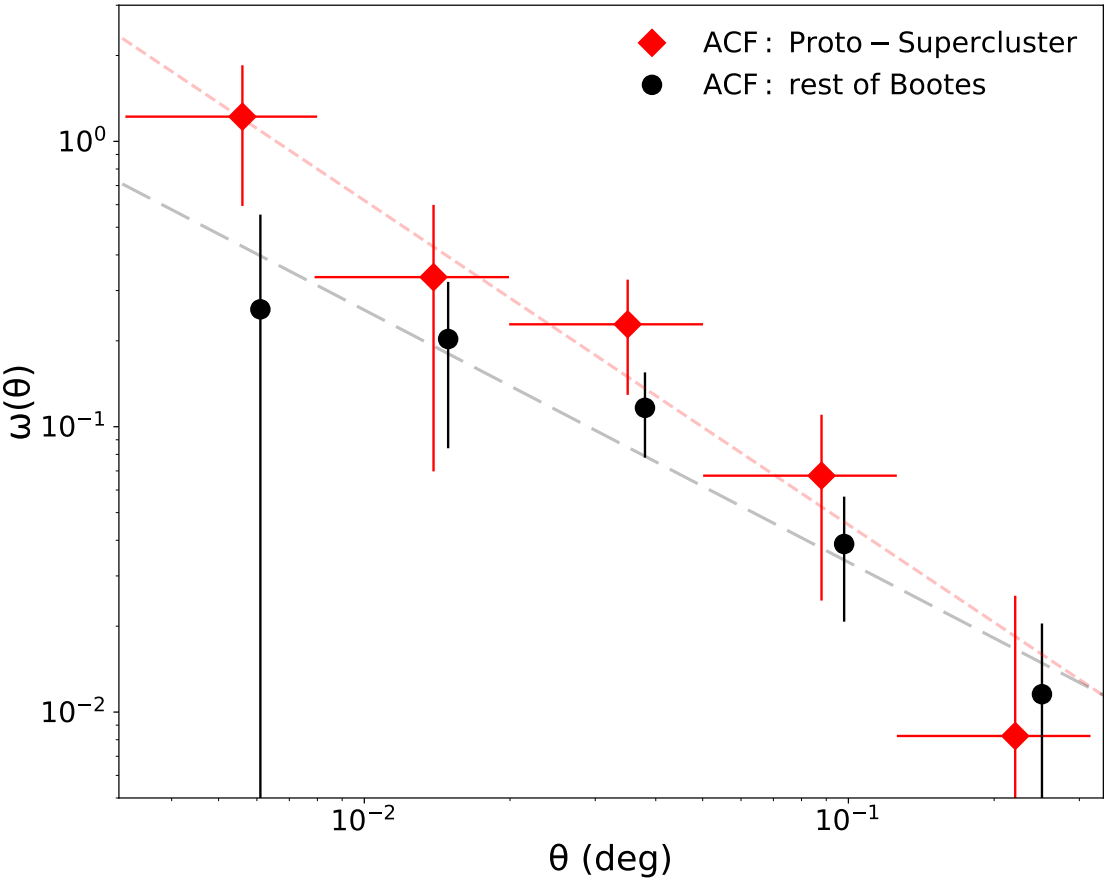

Figure 17: Two-point autocorrelation function (ACF) of the DOGs in the proto-supercluster region (red diamonds) and outside the proto-supercluster in Boötes (black circles). A power law of the form $\omega(\theta) = A_\omega \theta^{-\delta}$ was used to fit the observed data points. The dashed red (black) line represents the best-fit power-law for the observed ACF of the DOGs in the proto-supercluster (rest of the Boötes area).

which implies the presence of many clusters and protocluster nodes. The correlation function was modeled using a power-law of the form: $\omega(\theta) = A_\omega \theta^{-\delta}$. Here, A_ω and δ are the best-fit amplitude and slope of the power law, respectively. We obtained a slightly higher $\delta = 1.14^{+0.43}_{-0.32}$ for the DOGs in proto-supercluster region suggesting a stronger correlation. The corresponding value is $0.88^{+0.29}_{-0.28}$ for the rest of the Boötes field. However, in both cases, the values are consistent with previous clustering analysis of the DOGs in Boötes [17].

In summary, the DOG population presented in Dey et al. [34] was used to identify overdense regions of photo-z-selected galaxies in Boötes. As a result, a filamentary proto-supercluster comprising various galaxy overdensities, such as galaxy clusters and protoclusters, including the massive galaxy cluster IDCS J1426.5+3508 was identified at $z = 1.75$. We present the initial map of the filaments and substructures of the proto-supercluster. Follow-up multi-wavelength, deep photometric, and spectroscopic observations of the high-density nodes, the low-density filaments that connect the nodes will reveal a more accurate three-dimensional picture allowing the study of the gas kinematics, ICM dynamics, and the fate of the substructures across the complex proto-supercluster.

CHAPTER 3

ESTIMATING PHOTOMETRIC REDSHIFT PROBABILITY DISTRIBUTION FUNCTION USING GRADIENT BOOSTED REGRESSION TREE (GBRT)

Co-discipline Related Research Work

Published in *Euclid Collaboration et al. 2020, A&A, 644, A31*

DOI: 10.1051/0004-6361/202039403

Abstract

We report the performance of a machine-learning photometric redshift (photo-z) code for the upcoming *Euclid Space Mission*. We tested and implemented a decision tree-based architecture to compute the photometric redshift for a synthetic catalog with *Euclid*-like photometry as part of the *Euclid* Data Challenge 2 (DC2). For the *Euclid* sample, we obtained $\sigma_{\rm NMAD} = 0.057$ with an outlier fraction of $\sim 10\%$ in the field of COSMOS at $0.2 < z < 2.5$. This result, along with other photo-z code performances, indicates that none of the existing codes can achieve the desired accuracy of the Euclid mission. Therefore, it requires more homogeneous calibration datasets across a wide range of redshifts to obtain improved photo-z estimates.

3.1 Introduction

Upcoming wide-field surveys, such as the *Euclid Space Mission* [73] , *Vera C. Rubin Observatory* [64], *Nancy Grace Roman Space Telescope* [1] will produce

remarkably rich datasets of distant objects. These datasets will be exploited extensively for various extragalactic and cosmological studies. The prodigious amount of data will allow us to apply supervised machine learning (hereafter, ML) algorithms to unlock the hidden patterns in the datasets, and obtain predictions for various science parameters. For instance, obtaining photometric redshifts (photo-z) for millions of extragalactic sources using ML techniques has two major advantages: (i) computation time for ML redshifts is significantly less than the traditional template-fitting methods; (ii) ML redshifts are free from the biases arising from the high-z empirical templates that are not well-understood. In parallel with the progress in computational power over the past two decades, deep photometric and spectroscopic surveys have allowed exploring the redshift-photometry relationship via a host of ML techniques. For example, ANNz [27], and its new implementation [108] use neural networks for estimating redshift probability density functions (zPDF) using multi-wavelength photometry. Direct galaxy image stamps in different wavelengths are also used for extracting features and computing zPDFs using deep networks and decision trees [35, 56, 60, 96]. However, the main disadvantage of the supervised ML techniques is the dependency on training datasets with known accurate redshifts. ML methods performs only better when the training samples fully represent the galaxy photometric properties present in test samples [83]. A large sample of synthetic galaxies with *Euclid*-like photometry is thus ideal for testing various ML methods for cosmological applications.

In this project, we have independently developed and tested a ML-based

photo-z tool: a decision tree-based method for a synthetic catalog with *Euclid*-like photometry. We participated in the photometric redshift Data Challenge 2 (DC2) for the upcoming *Euclid* space mission. A total of 13 photometric redshift codes were tested by focusing primarily on the *Euclid* redshift range $0.2 < z < 2.6$ as part of the mission work. A variety of conventional template-fitting and ML codes were tested and evaluated using various metrics [41]. We tested and reported the performance of a decision tree-based machine learning tool called the Gradient Boosted Regression Tree (GBRT) [99].

3.2 Data

We used a DC2 calibration catalog to train the regression tree. For training purpose, we limited the sample by selecting only galaxies (including galaxies flagged as AGNs) with good quality spectroscopic redshifts and photometric redshifts where good spectroscopy is not available. Not all the galaxies in the calibration catalog has detection in all the DC2 bands (g, r, i, z, VIS, Y, J, H). We removed the galaxies from the training sample that do not have positive *signal-to-noise* ratio (SNR) in at least 4 of the DC2 bands. These prunings led to a training sample of $\sim 1.4 \times 10^5$ galaxies. For a robust training, we made 10 different brighter/fainter realizations of each galaxy. This resulted in a $10-$fold larger dataset ($\sim 1.4 \times 10^6$ galaxies) which was used ultimately in training the Gradient Boosted Regressor. There are total 8 filters in the DC2 catalog. We used flux values as the input features to train the data. We used $2''$ circular apertures for the g, r, i, z band fluxes while $1''$ circular

apertures were used for the Y, J, H band fluxes. For non-detection/non-observation of a galaxy in a filter, the flux value was set to zero and the corresponding error was set to a number representative of numerical infinity ($\sim 10^{30}$).

3.3 Analysis

3.3.1 Gradient Boosted Decision Tree (GBRT) Model

The Decision Tree Regressor is a simple tree regressor that divides the input space into n numbers of non-overlapping distinct regions. The number n is determined for which the error is minimized. For every observation that falls into a region R, the regressor makes same prediction for all of them, which is simply the mean of the response values of the training observation in the R space. The Gradient Boosted Decision Tree[1] is an algorithm that can be used for solving both regression and classification problems. GBRT starts off assuming an initial model that predicts the responses. Unlike other decision tree recipes, an additional model tries to fit the residuals — the difference between the actual response and predicted value — by attacking the weak parts of the residuals, and the initial model is then updated by adding the new learner that best fit the residuals. The role of the additional learner is to improve the shortcomings of the existing model. The residuals are computed again and another learner is computed that best fit the new residuals. This process continues until the loss function is minimized. Residuals are essentially negative gradients of the loss function. Therefore, a GBRT model is basically computed by

[1] `https://scikit-learn.org/stable/modules/ensemble.html`

iteratively updating the model by minimizing the gradient of the loss function.

3.3.2 Validation Catalog

A DC2 validation catalog was used to test the performance of the code. To calculate the probability distribution of the photometric redshift, P(z), of individual sources, we made 1000 slightly different realizations of each galaxy in the test catalog using a similar approach adopted for making multiple realizations for the training data. Prior to finalizing the sources for photo-z calculation, stars were removed using a gradient boosted classifier. Galaxies with detections in less than 4 bands were rejected as well. This led to a test catalog with $\sim 1.8 \times 10^5$ galaxies to check the photo-z code performance.

3.3.3 Star/Galaxy Classification

A gradient boosting classification algorithm was used to isolate the stars from the test data. The training sample contains $\sim 6\%$ stars while only $\sim 0.6\%$ of the training galaxies are flagged as AGNs. We limited the training only to classify the stars as the AGN fraction in the training data is too small to train the classifier. AGNs were therefore treated as normal galaxies. For training purpose, the classifier used the $8-$band flux values of the sources plus the VIS-band 'CLASS-STAR-DETECT' values from SExtractor as features. We tested the performance of the classifier by creating a validation catalog that consisted $\sim 2 \times 10^4$ randomly selected objects out of the training sample. The classification result is shown in Figure 18. The classifier identified 98.7% sources correctly in the validation catalog. We used

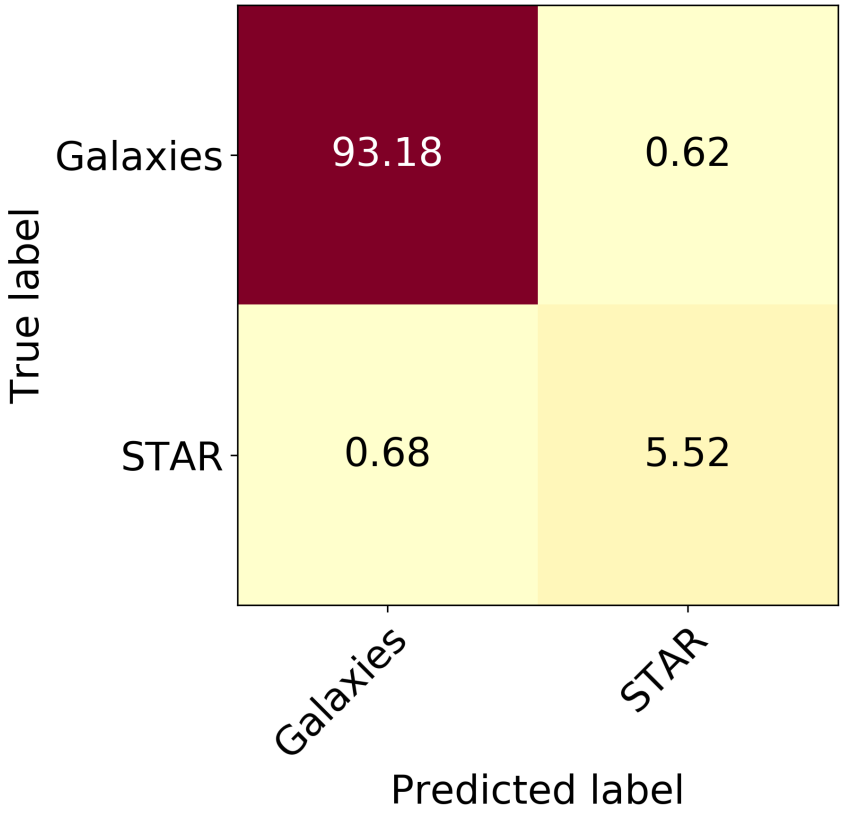

Figure 18: The classifier identified 98.7% of the sources correctly (the main diagonal of the confusion matrix). Only 1.3% sources were misidentified, of which 0.68% of sources (stars) were mislabelled as galaxies and 0.62% of sources (galaxies) were classified as stars. The validation sample, described in section 3.3.3 was used to generate this result.

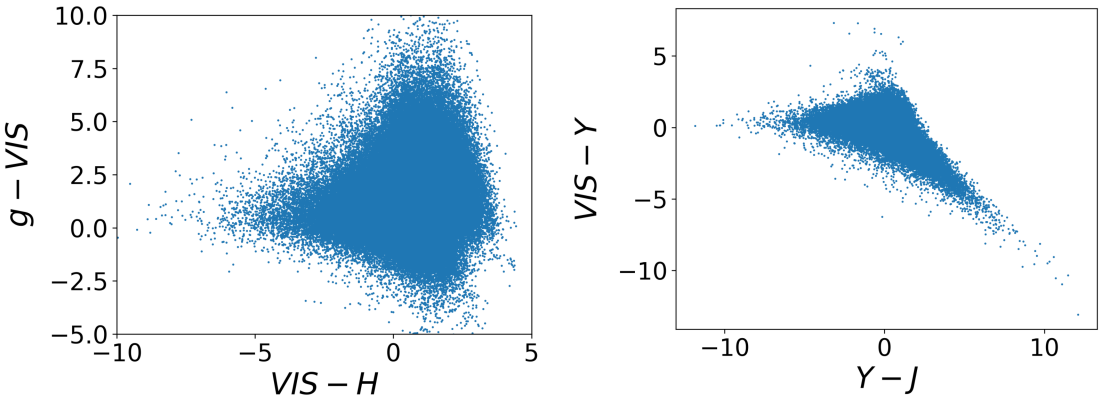

Figure 19: Left: The VIS-H color vs. g-VIS color of the training galaxies. Right: The Y-J color vs. VIS-Y color of the training galaxies.

the classifier to identify the fraction of stars in the test sample and $\sim 12,000$ sources were identified as stars which constitutes 6.22% of the test catalog.

3.3.4 Photo-z Reliability

The GBRT regressor found the g, VIS, Y, J, H band fluxes as the five most important features used in predicting the photo−z of the training galaxies. The regressor assigns importance to the features by calculating the fraction of galaxies they contribute to determine the photo−z. Figure 19 shows the distribution of the training galaxies in two different color-color spaces constituted by the g, VIS, Y, J, H-band magnitudes. If the colors of the test galaxies in these bands occupy the same space where the training colors are, then we define the photo−z as reliable.

3.3.5 Result

The probability distribution of the photometric redshifts, redshift point estimate, reliability of the photometric redshift and star/galaxy classification for the test catalog were submitted to *Euclid* OU-PHz unit as part of the data challenge. Figure 20 compares the prediction of the GBRT method to other machine-learning and template-fitting codes. GBRT performs well in comparison to other photometric redshift codes, such as La Phare and Phosphoros. However, GBRT's performance breaks down at $z > 2.5$ due to the lack of sufficient training sample.

3.4 Summary

Gradient boosted regression trees (GBRT) is a machine-learning method based on the sci-kit learn gradient boosted decision tree algorithm. For its training, galaxies and AGNs with good quality spectroscopic or 30-band photometric redshifts from L15 and detected in at least four bands were selected from the calibration catalog, leading to a training sample of around 1.4×10^5 sources. This sample size is increased by an order of magnitude by synthesizing 10 brighter and fainter versions of each source. The 2-FWHM aperture photometry in the g, r, i, and z bands, the 1-FWHM aperture photometry in the Y, J, and H bands, and the VIS total photometry were used to train the algorithm. A point estimate was determined for each source, and a probability distribution function was constructed via using 1000 realizations perturbed by a Gaussian error for each source. GBRT provides an indication of the most useful bands for the photo-z determination, those being the g,

Y, J, H, and VIS-like bands. Sources located in regions of this color space that were not covered by sources from the training sample were rejected. GBRT obtained a photometric redshift accuracy of 0.057(1+z) which is comparable to other standard photo-z computation techniques.

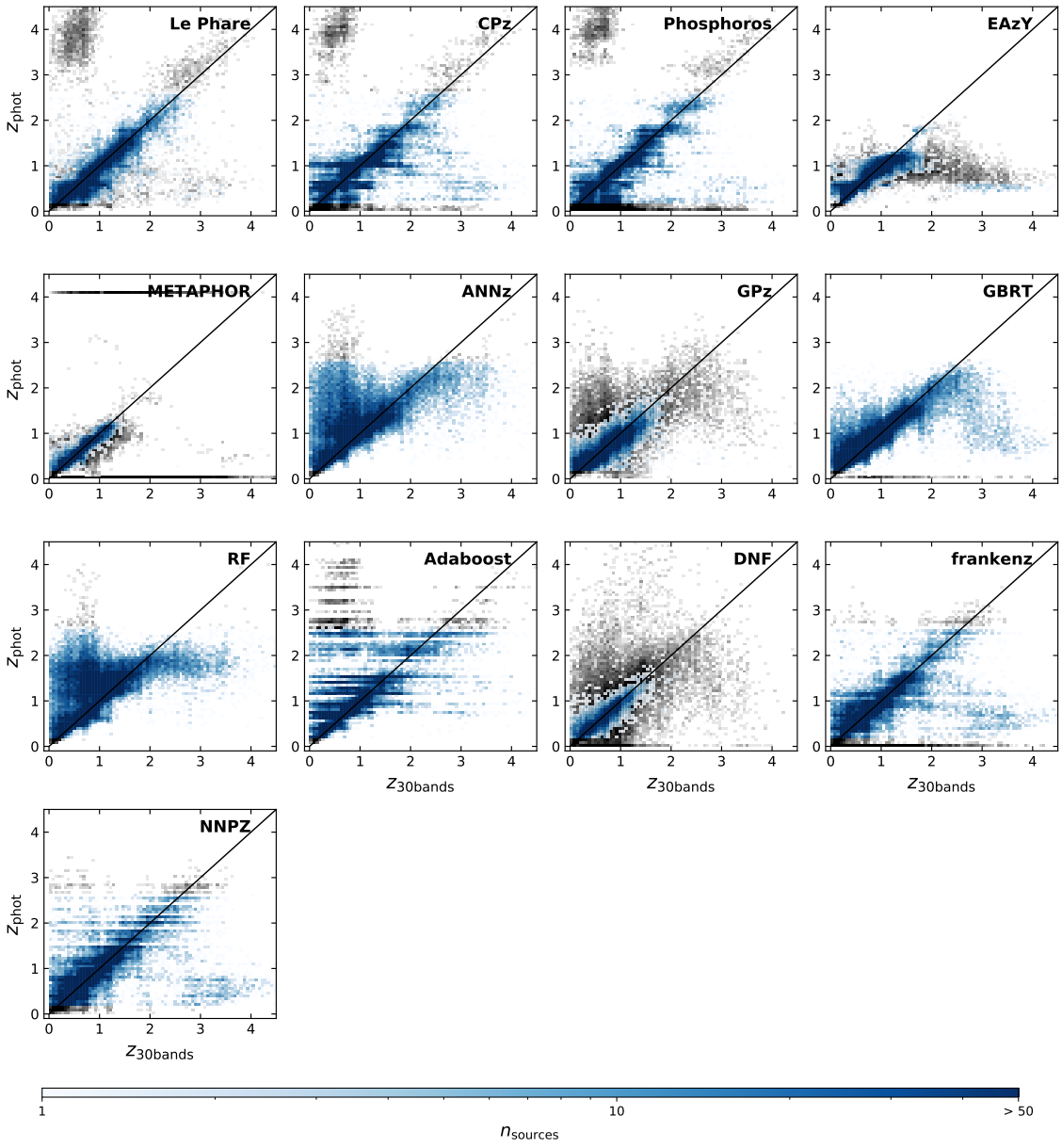

Figure 20: Density maps of the z_{phot} vs. z_{spec} are shown for all the methods tested in this study.

REFERENCE LIST

[1] Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv e-prints, arXiv:1902.05569.

[2] Alberts, S., Lee, K.-S., Pope, A., et al. 2021, MNRAS, 501, 1970,

[3] Ashby, M. L. N., Stern, D., Brodwin, M., et al. 2009, ApJ, 701, 428,

[4] Bahcall, N. A., Dong, F., Hao, L., et al. 2003, ApJ, 599, 814,

[5] Balogh, M. L., van der Burg, R. F. J., Muzzin, A., et al. 2021, MNRAS, 500, 358,

[6] Bayliss, M. B., Ashby, M. L. N., Ruel, J., et al. 2014, ApJ, 794, 12,

[7] Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559,

[8] Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57,

[9] Benson, B. A., Ade, P. A. R., Ahmed, Z., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII, ed. W. S. Holland & J. Zmuidzinas, 91531P,

[10] Bian, F., Fan, X., Jiang, L., et al. 2013, ApJ, 774, 28,

[11] Blanton, M. R., & Moustakas, J. 2009, ARA&A, 47, 159,

[12] Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, ApJS, 216, 27,

[13] Bleem, L. E., Bocquet, S., Stalder, B., et al. 2020, ApJS, 247, 25,

[14] Brodwin, M., Gonzalez, A. H., Moustakas, L. A., et al. 2007, ApJL, 671, L93,

[15] Brodwin, M., McDonald, M., Gonzalez, A. H., et al. 2016, ApJ, 817, 122,

[16] Brodwin, M., Brown, M. J. I., Ashby, M. L. N., et al. 2006, ApJ, 651, 791,

[17] Brodwin, M., Dey, A., Brown, M. J. I., et al. 2008, ApJL, 687, L65,

[18] Brodwin, M., Stern, D., Vikhlinin, A., et al. 2011, ApJ, 732, 33,

[19] Brodwin, M., Stanford, S. A., Gonzalez, A. H., et al. 2013, ApJ, 779, 138,

[20] Brown, M. J. I., Dey, A., Jannuzi, B. T., et al. 2007, ApJ, 654, 858,

[21] Casey, C. M. 2016, ApJ, 824, 36,

[22] Casey, C. M., Narayanan, D., & Cooray, A. 2014, , 541, 45,

[23] Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, ApJ, 779, 127,

[24] —. 2014, ApJL, 782, L3,

[25] Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. 2017, ApJL, 844, L23,

[26] Chung, S. M., Eisenhardt, P. R., Gonzalez, A. H., et al. 2011, ApJ, 743, 34,

[27] Collister, A. A., & Lahav, O. 2004, PASP, 116, 345,

[28] Cucciati, O., Zamorani, G., Lemaux, B. C., et al. 2014, A&A, 570, A16,

[29] Cucciati, O., Lemaux, B. C., Zamorani, G., et al. 2018, A&A, 619, A49,

[30] Dai, X., Assef, R. J., Kochanek, C. S., et al. 2009, ApJ, 697, 506,

[31] Dannerbauer, H., Kurk, J. D., De Breuck, C., et al. 2014, A&A, 570, A55,

[32] Decker, B., Brodwin, M., Saha, R., et al. 2021, arXiv e-prints, arXiv:2112.12239.

[33] Desai, V., Soifer, B. T., Dey, A., et al. 2008, ApJ, 679, 1204,

[34] Dey, A., Soifer, B. T., Desai, V., et al. 2008, ApJ, 677, 943,

[35] D'Isanto, A., & Polsterer, K. L. 2018, A&A, 609, A111,

[36] Dressler, A. 1980, ApJ, 236, 351,

[37] Duncan, K. J., Brown, M. J. I., Williams, W. L., et al. 2018, MNRAS, 473, 2655,

[38] Eisenhardt, P. R., Stern, D., Brodwin, M., et al. 2004, ApJS, 154, 48,

[39] Eisenhardt, P. R. M., Brodwin, M., Gonzalez, A. H., et al. 2008, ApJ, 684, 905,

[40] Estrada, J., Sefusatti, E., & Frieman, J. A. 2009, ApJ, 692, 265,

[41] Euclid Collaboration, Desprez, G., Paltani, S., et al. 2020, A&A, 644, A31,

[42] Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, in , Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1657–1669,

[43] Fassbender, R., Böhringer, H., Nastasi, A., et al. 2011, New Journal of Physics, 13, 125014,

[44] Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10,

[45] Galametz, A., Stern, D., De Breuck, C., et al. 2012, ApJ, 749, 169,

[46] Ghirardini, V., Bulbul, E., Kraft, R., et al. 2021, ApJ, 910, 14,

[47] Ghirardini, V., Bulbul, E., Hoang, D. N., et al. 2021, A&A, 647, A4,

[48] Giavalisco, M., Vanzella, E., Salimbeni, S., et al. 2011, ApJ, 743, 95,

[49] Gobat, R., Daddi, E., Onodera, M., et al. 2011, A&A, 526, A133,

[50] Gonzalez, A. H., Zaritsky, D., & Wechsler, R. H. 2002, ApJ, 571, 129,

[51] Gonzalez, A. H., Brodwin, M., Brown, M. J. I., et al. 2010, in American Astronomical Society Meeting Abstracts, Vol. 216, American Astronomical Society Meeting Abstracts #216, 415.13

[52] Gonzalez, A. H., Gettings, D. P., Brodwin, M., et al. 2019, ApJS, 240, 33,

[53] Hamilton, A. J. S. 1993, ApJ, 417, 19,

[54] Harikane, Y., Ouchi, M., Ono, Y., et al. 2019, ApJ, 883, 142,

[55] Hayashi, M., Koyama, Y., Kodama, T., et al. 2019, PASJ, 71, 112,

[56] Henghes, B., Thiyagalingam, J., Pettitt, C., Hey, T., & Lahav, O. 2022, MNRAS, 512, 1696,

[57] Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, ApJ, 696, 891,

[58] Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356,

[59] Houck, J. R., Soifer, B. T., Weedman, D., et al. 2005, ApJL, 622, L105,

[60] Hoyle, B. 2016, Astronomy and Computing, 16, 34,

[61] Huang, N., Bleem, L. E., Stalder, B., et al. 2020, AJ, 159, 110,

[62] Hwang, H. S., Park, C., Elbaz, D., & Choi, Y. Y. 2012, A&A, 538, A15,

[63] Imanishi, M., Nakagawa, T., Shirahata, M., Ohyama, Y., & Onaka, T. 2010, ApJ, 721, 1233,

[64] Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111,

[65] Jannuzi, B. T., & Dey, A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies, ed. R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner, 111

[66] Jiang, L., Wu, J., Bian, F., et al. 2018, Nature Astronomy, 2, 962,

[67] Jimeno, P., Broadhurst, T., Lazkoz, R., et al. 2017, MNRAS, 466, 2658,

[68] John, T. L. 1988, A&A, 193, 189

[69] Kochanek, C. S., Eisenstein, D. J., Cool, R. J., et al. 2012, ApJS, 200, 8,

[70] Kravtsov, A. V., & Borgani, S. 2012, ARA&A, 50, 353,

[71] Krishnan, C., Hatch, N. A., Almaini, O., et al. 2017, MNRAS, 470, 2170,

[72] Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64,

[73] Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193.

[74] Lee, S., & Park, C. 2002, Journal of Korean Astronomical Society, 35, 111,

[75] Lemaux, B. C., Cucciati, O., Tasca, L. A. M., et al. 2014, A&A, 572, A41,

[76] Limber, D. N. 1953, ApJ, 117, 134,

[77] Lotz, J. M., Papovich, C., Faber, S. M., et al. 2013, ApJ, 773, 154,

[78] Magliocchetti, M., Cirasuolo, M., McLure, R. J., et al. 2008, MNRAS, 383, 1131,

[79] Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2020, MNRAS, 496, 1554,

[80] Mantz, A. B., Abdulla, Z., Allen, S. W., et al. 2018, A&A, 620, A2,

[81] Martinache, C., Rettura, A., Dole, H., et al. 2018, A&A, 620, A198,

[82] Martini, P., Miller, E. D., Brodwin, M., et al. 2013, ApJ, 768, 1,

[83] Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, ApJ, 877, 81,

[84] Maturi, M., Bellagamba, F., Radovich, M., et al. 2019, MNRAS, 485, 498,

[85] McDonald, M., Allen, S. W., Bayliss, M., et al. 2017, ApJ, 843, 28,

[86] Muldrew, S. I., Hatch, N. A., & Cooke, E. A. 2015, MNRAS, 452, 2528,

[87] Murray, S. G., Diemer, B., Chen, Z., et al. 2021, Astronomy and Computing, 36, 100487,

[88] Murray, S. G., Power, C., & Robotham, A. S. G. 2013, Astronomy and Computing, 3, 23,

[89] Muzzin, A., Wilson, G., Demarco, R., et al. 2013, ApJ, 767, 39,

[90] Narayanan, D., Dey, A., Hayward, C. C., et al. 2010, MNRAS, 407, 1701,

[91] Newman, A. B., Ellis, R. S., Andreon, S., et al. 2014, ApJ, 788, 51,

[92] Noirot, G., Stern, D., Mei, S., et al. 2018, ApJ, 859, 38,

[93] Noordeh, E., Canning, R. E. A., King, A., et al. 2020, MNRAS, 498, 4095,

[94] Oguri, M., Lin, Y.-T., Lin, S.-C., et al. 2018, PASJ, 70, S20,

[95] Papovich, C. 2008, ApJ, 676, 206,

[96] Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26,

[97] Paterno-Mahler, R., Blanton, E. L., Brodwin, M., et al. 2017, ApJ, 844, 78,

[98] Peacock, J. A., & West, M. J. 1992, MNRAS, 259, 494,

[99] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine

Learning Research, 12, 2825

[100] Peebles, P. J. E. 1980, The large-scale structure of the universe

[101] Pimbblet, K. A., Shabala, S. S., Haines, C. P., Fraser-McKelvie, A., & Floyd, D. J. E. 2013, MNRAS, 429, 1827,

[102] Planck Collaboration, Aghanim, N., Altieri, B., et al. 2015, A&A, 582, A30,

[103] Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81,

[104] Rettura, A., Martinez-Manso, J., Stern, D., et al. 2014, ApJ, 797, 109,

[105] Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, ApJS, 154, 25,

[106] Roche, N., & Eales, S. A. 1999, MNRAS, 307, 703,

[107] Rosati, P., Stanford, S. A., Eisenhardt, P. R., et al. 1999, AJ, 118, 76,

[108] Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502,

[109] Santos, J. S., Altieri, B., Valtchanov, I., et al. 2015, MNRAS, 447, L65,

[110] Stanford, S. A., Eisenhardt, P. R., Brodwin, M., et al. 2005, ApJL, 634, L129,

[111] Stanford, S. A., Brodwin, M., Gonzalez, A. H., et al. 2012, ApJ, 753, 164,

[112] Steidel, C. C., Adelberger, K. L., Dickinson, M., et al. 1998, ApJ, 492, 428,

[113] Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, ApJ, 631, 163,

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

[115] Toshikawa, J., Malkan, M. A., Kashikawa, N., et al. 2020, ApJ, 888, 89,

[116] Toshikawa, J., Uchiyama, H., Kashikawa, N., et al. 2018, PASJ, 70, S12,

[117] Tran, K.-V. H., Nanayakkara, T., Yuan, T., et al. 2015, ApJ, 811, 28,

[118] Tully, R. B., Courtois, H., Hoffman, Y., & Pomarède, D. 2014, Nature, 513,

71,

[119] von der Linden, A., Wild, V., Kauffmann, G., White, S. D. M., & Weinmann, S. 2010, MNRAS, 404, 1231,

[120] Watson, C., Tran, K.-V., Tomczak, A., et al. 2019, ApJ, 874, 63,

[121] Weedman, D. W., Soifer, B. T., Hao, L., et al. 2006, ApJ, 651, 101,

[122] Willis, J. P., Clerc, N., Bremer, M. N., et al. 2013, MNRAS, 430, 134,

[123] Willis, J. P., Canning, R. E. A., Noordeh, E. S., et al. 2020, Nature, 577, 39,

[124] Wilson, G., Muzzin, A., Yee, H. K. C., et al. 2009, ApJ, 698, 1943,

[125] Wylezalek, D., Galametz, A., Stern, D., et al. 2013, ApJ, 769, 79,

[126] Wylezalek, D., Vernet, J., De Breuck, C., et al. 2014, ApJ, 786, 17,

[127] Younger, J. D., Bahcall, N. A., & Bode, P. 2005, ApJ, 622, 1,

[128] Zeimann, G. R., Stanford, S. A., Brodwin, M., et al. 2012, ApJ, 756, 115,

[129] —. 2013, ApJ, 779, 137,

VITA

Ripon Saha is a final-year Ph.D. student at the University of Missouri-Kansas City (UMKC). Mr. Saha joined the Masters in Physics Program at UMKC in 2014. In 2016, he joined the Ph.D. in Astrophysics program with Dr. Mark Brodwin at UMKC. Mr. Saha's primary research interest is to identify early-stage clusters of galaxies $\sim$10 billion light-years away in the young universe. Mr. Saha also develops and implements different machine-learning techniques to determine the distances to distant galaxies. Mr. Saha is currently leading three first-author projects, which will be published in reputed astronomy journals. Two of the journal papers are presently under review. He co-authored two other collaborative projects and presented at many major conferences and seminars. Mr. Saha is an active member of the upcoming joint ESA-NASA space program called the *Euclid Space Mission*. He will join Wells Fargo's eastern headquarter as a Quantitative Analytics Specialist in the upcoming summer.