

The eROSITA Final Equatorial-Depth Survey (eFEDS) Optical confirmation, redshifts, and properties of the cluster and group catalog[★]

M. Klein¹, M. Oguri^{2,3,4}, J. J. Mohr^{1,5}, S. Grandis¹, V. Ghirardini⁵, T. Liu⁵, A. Liu⁵, E. Bulbul⁵, J. Wolf^{5,19},
J. Comparat⁵, M. E. Ramos-Ceja⁵, J. Buchner⁵, I. Chiu^{6,7,8}, N. Clerc⁹, A. Merloni⁵, H. Miyatake^{4,10},
S. Miyazaki^{11,12}, N. Okabe^{13,14,15}, N. Ota^{16,17}, F. Pacaud¹⁶, M. Salvato⁵, and S. P. Driver¹⁸

(Affiliations can be found after the references)

Received 19 April 2021 / Accepted 14 June 2021

ABSTRACT

Context. In 2019, the eROSITA telescope on board the Russian-German satellite Spectrum-Roentgen-Gamma (SRG) began to perform a deep all-sky X-ray survey with the aim of identifying $\sim 100\,000$ clusters and groups over the course of four years. As part of its performance verification phase, a $\sim 140\text{ deg}^2$ survey, called eROSITA Final Equatorial-Depth Survey (eFEDS), was performed. With a depth typical of the all-sky survey after four years, it allows tests of tools and methods as well as improved predictions for the all-sky survey.

Aims. As part of this effort, a catalog of 542 X-ray selected galaxy group and cluster candidates was compiled. In this paper we present the optical follow-up, with the aim of providing redshifts and cluster confirmation for the full sample. Furthermore, we aim to provide additional information on the dynamical state, richness, and optical center of the clusters. Finally, we aim to evaluate the impact of optical cluster confirmation on the purity and completeness of the X-ray selected sample.

Methods. We used optical imaging data from the Hyper Suprime-Cam Subaru Strategic Program and from the Legacy Survey to identify optical counterparts to the X-ray detected cluster candidates. We make use of the multi-component matched filter cluster confirmation tool (MCMF), as well as of the optical cluster finder CAMIRA to derive cluster redshifts and richnesses. MCMF provided the probabilities with which an optical structure would be a chance superposition with the X-ray candidate. These probabilities were used to identify the best optical counterpart as well as to confirm an X-ray candidate as a cluster. The impact of this confirmation process on catalog purity and completeness was estimated using optical to X-ray scaling relations as well as simulations. The resulting catalog was furthermore matched with public group and cluster catalogs. Optical estimators of the cluster dynamical state were constructed based on density maps of the red-sequence galaxies at the cluster redshift.

Results. By providing redshift estimates for all 542 candidates, we construct an optically confirmed sample of 477 clusters and groups with a residual contamination of 6%. Of these, 470 (98.5%) are confirmed using MCMF, and 7 systems are added through cross-matching with spectroscopic group catalogs. Using observable-to-observable scaling and the applied confirmation threshold, we predict that 8 ± 2 real systems have been excluded with the MCMF cut required to build this low-contamination sample. This number agrees well with the 7 systems found through cross-matching that were not confirmed with MCMF. The predicted redshift and mass distribution of this catalog agree well with simulations. Thus, we expect that these 477 systems include $>99\%$ of all true clusters in the candidate list. Using an MCMF-independent method, we confirm that the catalog contamination of the confirmed subsample is $6 \pm 3\%$. Application of the same method to the full candidate list yields $17 \pm 3\%$, consistent with estimates coming from the fraction of confirmed systems of $\sim 17\%$ and with expectations from simulations of $\sim 20\%$. We also present a sample of merging cluster candidates based on the derived estimators of the cluster dynamical state.

Key words. catalogs – galaxies: clusters: general – galaxies: distances and redshifts – galaxies: clusters: intracluster medium – X-rays: galaxies: clusters

1. Introduction

Galaxy clusters are the most massive collapsed halos in the Universe. Their abundance is sensitive to cosmological parameters, which makes them valuable cosmological probes (e.g., Vikhlinin et al. 2009; Mantz et al. 2010; Rozo et al. 2010; Planck Collaboration XXIV 2016; Bocquet et al. 2019;

IDER Chitham et al. 2020; Abbott et al. 2020). Furthermore, galaxy clusters are exceptional astrophysics laboratories in which galaxy evolution, the dark matter self interaction cross section, cosmic ray acceleration and many other physical quantities can be studied (e.g., Dressler 1980; Moore et al. 1996; Clowe et al. 2006; van Weeren et al. 2010; Harvey et al. 2015).

An important topic for many cluster-related studies is understanding the cluster selection function and the purity of the cluster catalog. Cluster catalogs derived from X-ray observations have the advantage that the X-ray emission from galaxy clusters depends on the square of the electron density of the

[★] The catalog is only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/661/A4>

intracluster medium (ICM), which reduces the impact of projection effects of noncollapsed systems in the cluster catalog. X-ray surveys with a controlled selection function are therefore an excellent source for cluster-based studies (e.g., Rosati et al. 2002; Allen et al. 2011, for reviews). Previous works using either small-area deep data (Finoguenov et al. 2007; Leauthaud et al. 2010; Pacaud et al. 2016; Adami et al. 2018) or large but shallow surveys (Piffaretti et al. 2011; Klein et al. 2019; Finoguenov et al. 2020) have produced useful cluster catalogs, some including as many as a few thousand X-ray selected clusters. The largest number of X-ray clusters were found using the shallow but large-area ROSAT all-sky survey (RASS; Truemper 1982; Voges et al. 1999).

With eROSITA (Predehl et al. 2021; Merloni et al. 2012), the next-generation X-ray survey telescope recently started its operation and its journey to produce a new high-quality all-sky X-ray survey. As part of its performance verification program, a medium-area (~ 140 square degrees) survey was performed prior to the start of the all-sky survey. The average exposure time of ~ 1.3 ks after vignetting corrections is comparable to the time that will be reached in the final all-sky survey in the equatorial regions. The field location was chosen to lie within surveys with deep multiband-imaging data from the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018a) and the Legacy Survey (LS; Dey et al. 2019). It further overlaps the Sloan Digital Sky Survey (SDSS; Blanton et al. 2017) and one of the Galaxy And Mass Assembly (GAMA; Driver et al. 2009) fields, both of which provide significant amounts of spectroscopic data. This makes the eROSITA final equatorial depth survey (eFEDS) a good and early test case for validating and improving the tools and methods needed for the scientific exploration of the all-sky survey.

This paper is part of the eROSITA early data release and focuses on the optical confirmation and redshift estimates of X-ray selected cluster and group candidates described in Liu et al. (2022a). The goal is to provide information for a clean and reliable cluster catalog in which contamination and incompleteness due to optical cleaning are clearly understood. We also provide optical centers, estimators of the cluster dynamical state, and richness measurements as an additional mass estimator for each cluster.

2. Datasets

We restrict the description of datasets to the three main resources used in this work: the X-ray data from eROSITA, and the optical imaging surveys HSC-SSP and LS. Each of these are described in dedicated papers on the data processing and data products. We therefore only recall the most crucial aspects needed for this paper. The eFEDS footprint and its coverage with optical data is shown in Fig. 1.

2.1. X-ray dataset

The extended ROentgen Survey with an Imaging Telescope Array (Predehl et al. 2021) on board the Spectrum-Roentgen-Gamma (SRG) consists of seven telescope modules (TMs), acting as seven separate telescopes observing the same circular patch of the sky with a diameter of ~ 1 degree. The field-of-view average angular resolution is $\sim 26''$ (HEW at 1.49 keV) and $\sim 18''$ on-axis (Predehl et al. 2021), and its effective area in the 0.5–2 keV band of the seven TMs is similar to that of the combined PN+MOS instruments on *XMM-Newton*.

The eFEDS field was observed by eROSITA in the first quarter of November 2019 as part of the performance verification phase (Brunner et al. 2022). The total time spent on this field is ~ 350 ks, corresponding to an average exposure time of ~ 2.3 ks and ~ 1.3 ks after accounting for vignetting. The data were processed with the eROSITA Standard Analysis Software System (eSASS; Brunner et al. 2022).

The X-ray cluster candidate list contains 542 sources, selected by requiring extent likelihoods greater 6, a detection likelihood greater than 5, and a minimum source extent of 2 pixels ($8''$). From extensive X-ray simulations of the eFEDS field (Comparat et al. 2020), we expect that $\sim 80.3\%$ of these sources correspond to genuine galaxy clusters. For more details, we refer to the dedicated paper on the X-ray catalogs (Brunner et al. 2022; Liu et al. 2022b).

2.2. Hyper Suprime-Cam Subaru Strategic Program

The Hyper Suprime-Cam Subaru Strategic Program (Aihara et al. 2018a) is an imaging survey conducted using the Hyper Suprime-Cam (HSC) camera on the 8.2 m Subaru telescope at Mauna Kea, Hawaii. HSC is a wide-field (1.7 degree diameter) optical imager (Miyazaki et al. 2018) installed at the prime focus of the Subaru telescope. The HSC-SSP survey consists of three different layers called Wide, Deep, and Ultradeep. Of interest here is the Wide layer, which is conducted in five broad bands (*grizy*) over a total area of 1400 square degrees to 5σ depths of about 26.6, 26.2, 26.2, 25.3, and 24.5. Details of the data processing and source detection can be found in Bosch et al. (2018), and Aihara et al. (2018b, 2019).

The most recent S20A data from HSC-SSP yield a coverage of $\sim 95\%$ of the eFEDS footprint in *griz*-band. In contrast, the public data release 2 (PDR2; Aihara et al. 2019) only contains data out to S18A (2018). We therefore use the non-public data stemming from observations until 2019 and 2020, which are referred to as S19A and S20A, respectively. Throughout the paper, we use *cmode1* magnitudes derived from light profile fitting as total magnitudes of galaxies, while the colors of each galaxy are derived from the point spread function (PSF)-matched aperture photometry without deblending with the target PSF FWHM of $1''.3$ and the aperture diameter of $1''.5$ in order to mitigate the issue of deblending failure in crowded areas (Oguri et al. 2018; Aihara et al. 2018b). We only use galaxies with *z*-band *cmode1* magnitudes after correction for Galactic extinction (Schlegel et al. 1998) brighter than 24, which corresponds to a 10σ detection significance for extended sources, and their errors smaller than 0.1 mag. We select extended objects based on the star-galaxy separation in *i*-band images (*i_extendedness_value*) because *i*-band images tend to be taken in better seeing conditions for weak-lensing shape measurements.

2.3. Legacy Survey

The DESI Legacy Imaging Surveys (LS; Dey et al. 2019) is a combination of four imaging surveys: the 9000 deg^2 *grz*-band DECam-based DECaLS survey; the 5000 deg^2 BASS and MzLS surveys, which provide photometry in *gr* and *z* band, respectively; and the WISE and NEOWISE surveys in the mid-IR at 3.4 and 4.6 μm . Data release 8 of LS further includes archival observations and covers an area of 19000 deg^2 in total. Source detection and photometry is performed using the tractor¹

¹ <http://ascl.net/1604.008>

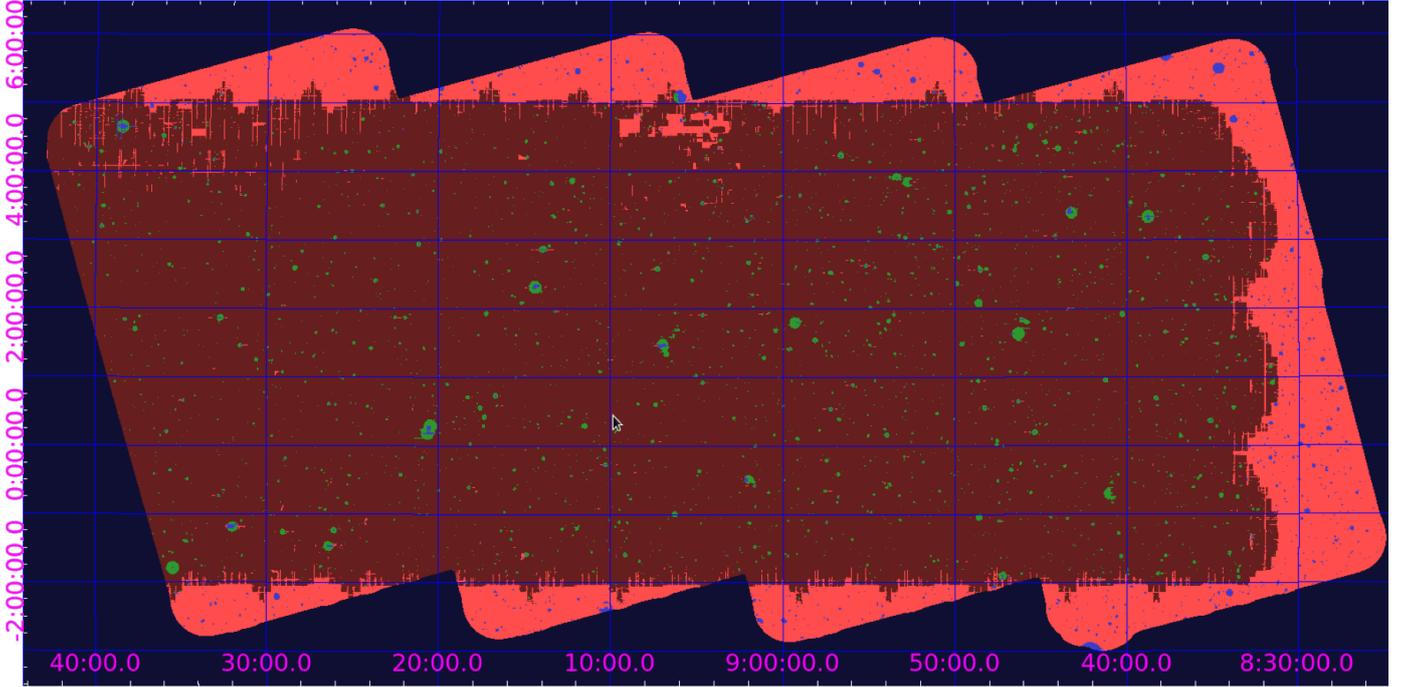


Fig. 1. eFEDS footprint and its coverage with HSC and LS data. Dark red shows the HSC and LS coverage, and light red shows the area that is covered by LS alone. Green and blue indicate masked regions. The masking is mostly caused by bright star masks. In addition to the masked regions, the full eFEDS footprint is covered either by HSC or LS.

software (Lang et al. 2016) in optical bands. On the mid-IR data, forced photometry with deblending is performed to take the PSF into account. The eFEDS footprint lies within the DECaLS footprint, which shows median 5σ depths of 23.72, 23.27, and 22.22 in grz band.

3. Methods

To define the cluster and group candidate catalog, a compromise has to be made between catalog purity and completeness, and related to this, catalog size. By examining simulated X-ray surveys of eFEDS-like fields (Brunner et al. 2022; Liu et al. 2022b) that used the same definition for the candidate list, we expect about $\sim 19.3\%$ of the cluster candidates to be noncluster contaminants. Furthermore, the X-ray selection function strongly depends on redshift, reaching mass ranges well within the group regime at low redshifts while probing relatively high-mass systems at the high-redshift end.

The optical follow-up therefore has to overcome two challenges: (1) identifying systems that are likely not real clusters at all, and (2) assigning the best optical counterpart when multiple systems lie along the line of sight. To this end, we have developed the multicomponent matched filter (MCMF) cluster confirmation tool (Klein et al. 2018, 2019). In addition to applying MCMF, we execute a forced run of the CAMIRA optical cluster finder (Oguri 2014; Oguri et al. 2018) and then cross match to the standard CAMIRA optical cluster catalog using optical data from the HSC-SSP.

3.1. MCMF

The MCMF method is described in detail elsewhere (Klein et al. 2018, 2019). In this section we therefore restrict ourselves to the basic description and to modifications compared to previous

work. The core of MCMF makes use of the red sequence of cluster galaxies (Gladders & Yee 2000) and calculates the weighted number, called richness (λ), of excess galaxies within a certain magnitude and radial range around the X-ray position.

MCMF calculates the richness in 230 redshift bins out to $z = 1.3$. For each redshift bin an aperture corresponding approximately to r_{500} is estimated based on the X-ray count rate, the redshift bin, and an observable-mass scaling relation.

We used the luminosity-mass scaling relation given in Bulbul et al. (2019),

$$L_{500,0.5-2.0 \text{ keV}} = A_X \left(\frac{M_{500}}{M_{\text{piv}}} \right)^{B_X} \left(\frac{E(z)}{E(z_{\text{piv}})} \right)^2 \left(\frac{1+z}{1+z_{\text{piv}}} \right)^{\gamma_X}, \quad (1)$$

where A_X , B_X , and γ_X were found to have best-fit values of $4.15 \times 10^{44} \text{ erg s}^{-1}$, 1.91, and 0.252, respectively. The pivot mass M_{piv} and redshift z_{piv} were $6.35 \times 10^{14} M_{\odot}$ and 0.45. For the richness estimate we solved this equation for M_{500} to obtain r_{500} using a simplified conversion of count rate to luminosity. For this conversion we took the count rate from the source detection and assumed a MEKAL model with a fixed temperature of 3 keV, a metallicity of $0.3 Z_{\odot}$, and cosmology of $\Omega_M = 0.3$, $\Omega_{\Lambda} = 0.7$, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ to obtain an approximate L_{500} . With this, the conversion from count rate to our approximate M_{500} solely depends on redshift and hydrogen column density, where the latter is unimportant for the column densities found in the eFEDS area.

The distribution of richness as a function of redshift obtained in this way was then searched for peaks, and up to three potential candidates were registered. The same approach was then repeated with randomized positions that exclude regions around the real cluster candidates, which we will call randoms throughout the paper. With the set of richnesses and redshifts of real candidates and randoms, we calculated for each candidate i the

estimator $f_{\text{cont},i}$, which is defined as

$$f_{\text{cont},i} = \frac{\int_{\lambda_i}^{\infty} f_{\text{rand}}(\lambda, z_i) d\lambda}{\int_{\lambda_i}^{\infty} f_{\text{obs}}(\lambda, z_i) d\lambda}, \quad (2)$$

where $f_{\text{rand},z}$ is the richness distribution of randoms at the cluster candidate redshift z_i , $f_{\text{obs}}(\lambda, z_i)$ is the richness distribution of true candidates, and λ_i is the richness of the cluster candidate. Because the numbers of candidates and independent randoms are limited, the calculations were performed in multiple redshift bins, and estimates for specific redshifts were then derived using interpolation.

The estimator $f_{\text{cont},i}$ is correlated with the probability of a source to be a chance superposition, and it can therefore be used to control the overall sample contamination. A selection of all sources $f_{\text{cont},i} < 0.3$ means that we statistically allow for a 30% contamination, caused by sources showing similar richnesses and redshifts as the candidates that nonetheless have no physically associated extent-selected X-ray source. These are random superpositions of X-ray candidates with optical systems that just happen to lie along the line of sight. As long as the contaminants allowed by the optical selection and the contaminants allowed by the X-ray selection are uncorrelated, the applied cut in f_{cont} can be seen as a factor describing the reduction of the initial contamination of the cluster candidate catalog. Furthermore, the estimate of f_{cont} for each of the peaks found for a single cluster candidate allows selecting the most likely counterpart as the one showing the lowest f_{cont} . Ambiguity only arises if multiple sources show very low f_{cont} . This is typically the case for $\sim 2\%$ of the sources and can be traced by the difference between the lowest and the second lowest f_{cont} estimate of a given source.

We note here that all X-ray masses called approximate M_{500} in this paper refer to the estimates used for the richness measurement. A bias as well as redshift and mass trends to true halo mass can be expected. This can impact the estimated richness, but has no significant impact on the redshift estimation. Because f_{cont} is derived as a function of redshift and using randoms that share the same characteristics as the candidates, the estimate is robust against systematics in the adopted relation of M_{500} versus halo mass. We also note here that in parallel to this work, X-ray count rates and luminosities are re-extracted and measured within r_{500} (Bahar et al. (2022)). Additionally, a new luminosity-mass relation is obtained using HSC weak lensing (Chiu et al. 2022).

3.1.1. MCMF on HSC

Similar to our previous work using data from the Dark Energy Survey (DES), we used the *griz* bands. We omitted the *y* band because it does not significantly contribute to the MCMF performance. This choice also allowed us to maximize the footprint with full color coverage. We used clusters with spectroscopic redshifts over the full HSC-SSP area to calibrate λ versus redshift peak profiles and the red-sequence models. The HSC flagging to mask regions around bright stars was chosen quite conservatively due to drivers from studies of weak gravitation lensing (Coupon et al. 2018), and therefore the masked regions cover a significant fraction of the footprint. Originally, we performed two MCMF runs, one with and one without the flags. We did not find significant issues with the run ignoring the near bright source flag and therefore decided to use this setting as default. However, we provide a flag when a source center lies within a flagged region. We also report the fraction of area lies within r_{500} that is flagged.

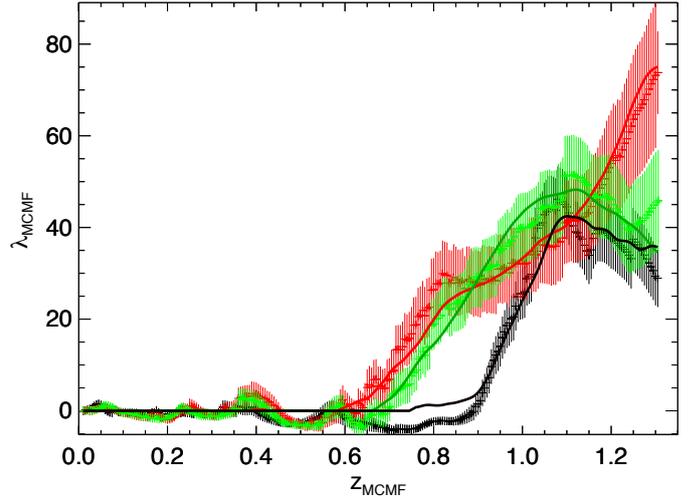


Fig. 2. MCMF richness vs. redshift plot for the $z = 1.1$ cluster eFEDS J084044.7+024108. In red and green we plot the output of MCMF on LS data, using *grz* and *grzW1*, respectively. The run on HSC is shown in black. The continuous lines show the best-fit peak profile at the cluster redshift.

3.1.2. MCMF on Legacy Survey

For the Legacy Survey we used the *grz* bands and the W1 band from unWISE (Lang 2014; Meisner et al. 2019). MCMF was run twice: once with *grz* band, and once with the W1 band as well. The *z*-W1 color provides improved redshifts at the high-redshift end, and the tractor photometry addresses the significant blending that occurs in the center of clusters in the W1 bands. The star-galaxy separation was performed out to $z = 21.7$ mag, and the residual stellar contamination was removed statistically using a local background for the richness measurement.

An example of the redshift estimate of the $z = 1.1$ cluster eFEDS J084044.7+024108 is shown in Fig. 2. MCMF fits so-called peak profiles to the distribution of λ versus redshift to obtain best-fit redshift and richness estimates. MCMF finds consistent redshifts in all three MCMF runs. The MCMF runs on the different optical data resulted in vastly different peak shapes. Nonetheless, the predicted peak shapes and the observed peaks for each optical dataset agree well. At the high-redshift end, the peak profile changes significantly from a steady rising function to a peaked profile when the *w1* band is added to the LS MCMF. This shows that the *w1*-band data help improve the MCMF performance at high redshift.

3.1.3. Combining HSC and LS results

Similar to what is shown in Fig. 2, the structures found in the MCMF runs on the different datasets typically agree excellently with one other. Differences first occur at the high-redshift end where the shallower LS data and the missing *i*-band information causes redshift and richness estimates to be noisier. A second reason for mismatches is the local lack of usable data either due to bright star masks or patchy data. A third obstacle is that the S19A data, and even more frequently, S20A HSC data, suffer from photometric zeropoint issues at a few locations (see Aihara et al. 2019). MCMF combines the red-sequence-based weights of all three colors, $g - r$, $r - i$, and $i - z$. If at least one color is significantly different in terms of the expected width of the red sequence, the overall richness is significantly biased low. Because the observed red-sequence width depends on redshift,

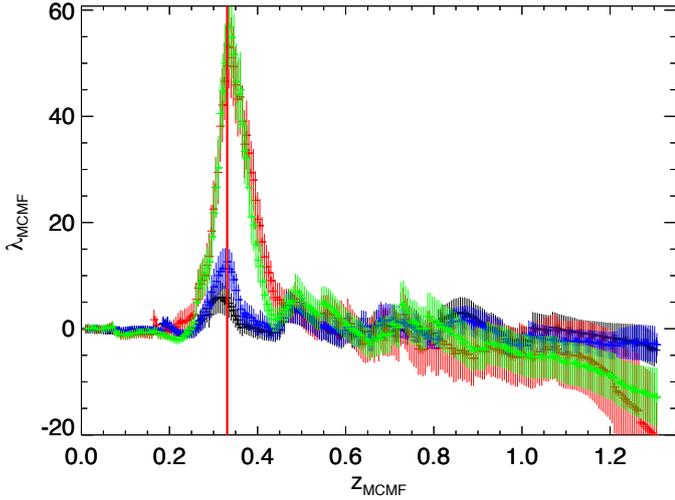


Fig. 3. Similar to Fig. 2, richness vs. redshift plot for eFEDS J092739.7-010427 for four MCMF runs. The runs using LS (red and green) both show a clear cluster peak and agree well with one other. The MCMF runs using HSC S19A data (black) and S20A data (blue) show a significantly smaller peak due to local offsets in the photometric zeropoint.

the impact of the zeropoint issues is most prominent at low redshifts. An example for this effect is shown in Fig. 3.

As the LS suffers less from this issue and provides homogeneous coverage over the full eFEDS footprint, we chose the MCMF run on LS in the *grzW1* mode as default for cluster confirmation for $z < 0.5$. For counterparts at higher redshifts, the HSC measurements were used. For sources that lack local data in either of the surveys, the entries from the least affected MCMF run were used.

The combined redshift, richness, and confirmation estimates are provided alongside the key entries of the MCMF runs on both surveys in the results table. The best counterpart was chosen as the counterpart with the lowest f_{cont} estimate.

3.1.4. Spectroscopic redshifts

The location of the eFEDS field was partially chosen because it overlaps spectroscopic surveys such as the 2MRS (Huchra et al. 2012), SDSS (Blanton et al. 2017), and GAMA (Driver et al. 2009). We used the public data and list spectroscopic redshifts for clusters that either have a measurement for the brightest cluster galaxy (BCG) or for at least three cluster members within r_{500} .

3.1.5. Optical estimators of the cluster dynamical state

As in Klein et al. (2019), we calculated estimators of the cluster dynamical state based on the optical data. The estimators in this work are mostly identical to those computed in our previous work. We therefore only briefly summarize the provided estimators.

A first set of three estimators is closely related to the estimators described in Wen & Han (2013). The main difference between Wen & Han (2013) and the estimators provided here is the different type of galaxy density map. Wen & Han (2013) used maps for the dynamical state estimators that were based on galaxy positions and r -band luminosities of sources with redshifts within 4% of the cluster redshift. In this work we used the galaxy density maps from the MCMF pipeline, which includes

galaxies consistent with the red sequence at the cluster redshift smoothed with a fixed 125 kpc Gaussian kernel.

The asymmetry factor α is defined as the ratio of the difference power over the total fluctuation power within r_{500} ,

$$\alpha = \frac{\sum_{i,j} [I(x_i, y_j) - I(-x_i, -y_j)]^2 / 2}{\sum_{i,j} I^2(x_i, y_j)}, \quad (3)$$

where $I(x_i, y_j)$ is the value of the density map at the cluster-centric position x_i, y_j . The normalized deviation δ is based on a fit of a 2D King model (King 1962),

$$I_{2\text{Dmodel}}(x, y) = \frac{I_0}{1 + (r_{\text{iso}}/r_0)^2}, \quad (4)$$

where I_0 is the peak intensity at the cluster center, r_0 is the characteristic radius, and r_{iso} describes the profile isophote with $r_{\text{iso}}^2 = (x \cos \theta + y \sin \theta)^2 + \epsilon(-x \sin \theta + y \cos \theta)^2$. The estimator δ is then the normalized deviation of the residual map within r_{500} after subtraction of the 2D model,

$$\delta = \frac{\sum_{i,j} [I(x_i, y_j) - I_{2\text{Dmodel}}(x_i, y_j)]^2}{\sum_{i,j} I^2(x_i, y_j)}. \quad (5)$$

The last estimator is the ridge flatness β and is derived by fitting a 1D king profile $I_{1\text{D}} = I_0 / (1 + (r/r_0)^2)$ to different sectors of the galaxy density map. Using the concentration estimator c_{King} as $c_{\text{King}} = r_{500}/r_0$, we searched for the lowest concentration out of thirty-six 10° wide angular wedges centered on the cluster. We call this the concentration of the ridge $c_{\text{King,R}}$. The ridge flatness is then defined with respect to the median of the derived concentrations as

$$\hat{\beta} = \frac{c_{\text{King,R}}}{\tilde{c}_{\text{King}}}. \quad (6)$$

To ensure positive scaling with the other estimators, we simply redefined the original estimator in Wen & Han (2013) to $\beta = 1 - \hat{\beta}$, while $\hat{\beta}$ corresponds to the original definition.

While Wen & Han (2013) based their study on optically selected clusters and thus always started at the optical center, our work starts with X-ray selected clusters and therefore X-ray centers. To obtain the estimator described above, we therefore left the 2D profile free to recenter. When we assume that the X-ray center provides a good estimate of the position of the ICM and the optical center that of the dark matter, then the offset between both centers yields valuable information about the cluster dynamical state. In practice, offsets are not only driven by the gas to halo center offsets, but by the complex shapes of the galaxy distributions, which in turn are again correlated with cluster dynamical state. We therefore list the center offset in terms of approximate r_{500} as an additional proxy of the dynamical state.

The estimators so far either probe the disagreement between the model and data (δ) or the asymmetry (α, β). Systems that are symmetric and well described by the model are therefore down-weighted by these estimators. Given the freedom of the 2D model and the resolution of the smoothed map, merging systems may be decently fit by the model, but with unusual fit parameters. In addition to the offset between 2D model center and X-ray position, the estimator that is most sensitive to merging is the 2D profile ellipticity. High ellipticities found by the model could either be related to very elliptical halos, which might arise from a cluster merger event, or to the code fitting two merging clusters with one model centered between both systems.

All these estimators are based on the fit of profiles to the galaxy density map. Therefore they strongly share systematics related to the fit of the data. Furthermore, the interest may preferentially be in mergers with at least two pronounced overdensities. To address this, we used the galaxy density map and the source-detection tool SExtractor (Bertin & Arnouts 1996) to identify overdensities in the map. We selected the nearest overdensity that has a $FLUX_ISO$ measurement of at least 25% of that overdensity, which is assumed to be the counterpart of the X-ray source. The $FLUX_ISO$ measurement of SExtractor can be interpreted in this context as a noisy richness estimate that should scale with the mass of the structure. For all cluster candidates reaching this threshold, we list the $FLUX_ISO$ ratio and the offset distance in units of r_{500} of the main cluster to the nearest such overdensity. As a last information for selecting merging systems, we also provide the approximate X-ray mass and distance of the nearest neighbor in our MCMF-confirmed cluster catalog. This can be seen as the analog to the optically based estimator. While the optical estimator may fail due to noise and projections of nearby noncollapsed systems, the X-ray estimator alone might be triggered by source splitting or projection of AGN flux into a relaxed system. The combination of the two estimators should produce a rather clean sample of multiple systems within a certain radius, however. This is especially true if both estimators point to the same neighboring system, as indicated through similar values for the nearest neighbor.

3.2. CAMIRA

CAMIRA is a cluster-finding algorithm based on overdensities of red-sequence galaxies (Oguri 2014). Multiband galaxy colors are fit with the stellar population synthesis model of Bruzual & Charlot (2003) with a red-sequence galaxy template as a function of redshift. Specifically, the formation redshift of galaxies is fixed to $z = 3$, and the stellar mass dependence of the metallicity is included to reproduce the tilt of the red sequence. To improve the accuracy of the model, corrections of colors that are derived by fitting the model to galaxies with spectroscopic redshifts are included. The likelihood of fitting as a function of redshift is converted into a number parameter, based on which the 3D richness map is constructed using a spatial filter with a transverse distance of ~ 1 Mpc. Cluster candidates are selected from peaks in the richness map. A massive red-sequence galaxy near each peak is selected as the BCG of the cluster, and richness and photometric redshift (photo- z) estimates are repeated around the BCG. For more details of the CAMIRA algorithm, we refer to Oguri (2014).

Oguri et al. (2018) applied the CAMIRA algorithm to the HSC-SSP S16A dataset, which covers ~ 230 deg², to construct a catalog of ~ 1900 clusters at redshift $0.1 < z < 1.1$ and richness $N > 15$. The comparison with spec- z s of the BCGs indicates that the cluster photo- z s are accurate. The bias and scatter in $\Delta z / (1 + z)$ are better than 0.005 and 0.01, respectively, over a wide redshift range. A careful weak-lensing analysis of these CAMIRA HSC-SSP clusters by Murata et al. (2019) suggests that the richness correlates with the halo mass with a relatively low scatter and that a richness threshold of 15 corresponds to a halo mass threshold of $\sim 10^{14} h^{-1} M_{\odot}$ (see also Chiu et al. 2020b,a). The richness and mass relation shows a tight correlation regardless of dynamical states, in contrast to the Y_{SZ} -mass and X-ray luminosity-mass relations (Okabe et al. 2019).

While the default setup of CAMIRA finds clusters of galaxies around peaks of the 3D richness map, CAMIRA can also find clusters of galaxies around positions provided by external

catalogs. In this mode, peaks of the 3D richness map are simply replaced by the positions of the external catalogs. CAMIRA then scans the richness at these positions as a function of redshift to identify cluster candidates, and searches the BCGs around the cluster candidates. This forced CAMIRA algorithm has successfully been applied to the ACT-DR5 Sunyaev-Zel'dovich (SZ) cluster catalog (Hilton et al. 2021) and the XXL X-ray cluster catalog (Willis et al. 2021) to construct CAMIRA cluster catalogs around the SZ and X-ray peaks, respectively.

4. Application and results

4.1. Comparing CAMIRA and MCMF results

The CAMIRA and MCMF codes both make use of the red sequence of galaxies in galaxy clusters and its dependence on redshift. Therefore we expect similar performance in key properties such as redshift and richness estimates. As discussed in Sect. 3.2, CAMIRA can run in two modes: one as a classical cluster finder, and the other using the X-ray position as prior. We adopted two CAMIRA cluster catalogs constructed using these two modes for our analysis. We compared the MCMF results with both CAMIRA runs because one run can provide insights into the optical cluster-finding algorithm while the other run can maximize the information about the X-ray cluster candidates.

For the match with the normal catalog, we used a maximum offset of 2 arcminutes and kept the nearest match to the X-ray position. We found 239 matches. All but three of these are $f_{\text{cont}} < 0.3$ and therefore assumed to be confirmed. The three remaining sources have similar redshifts in both MCMF and CAMIRA, but have a low richness and therefore are likely not a proper counterpart of the X-ray source. The redshifts of the 236 $f_{\text{cont}} < 0.3$ found by both codes are shown in Fig. 4. We investigate the seven most extreme photo- z outliers, which are shown in yellow in Fig. 4. We find for all these systems that the second-ranked MCMF system is consistent with the system matched using CAMIRA and that the first-ranked MCMF system is the better match to the X-ray source. These systems have a low richnesses and are therefore likely fall below the cut of 15 member galaxies in the normal CAMIRA catalog.

In case of the forced CAMIRA run, we find 324 sources with richness $\lambda > 10$ and 316 that show $f_{\text{cont}} < 0.3$. Again, the $f_{\text{cont}} > 0.3$ systems with CAMIRA matches are also found by MCMF either as best or second best counterpart. The situation for $f_{\text{cont}} < 0.3$ systems is also similar to the normal CAMIRA run; MCMF usually also finds the CAMIRA systems, but ranks another, typically lower redshift, cluster as the better counterpart. This is because with f_{cont} , MCMF takes the redshift-dependent X-ray selection into account.

In the third panel of Fig. 4 we show the richness comparison between CAMIRA and MCMF. In addition to scatter between the two estimates, a divergence from the one-to-one scaling is visible. This can be explained by the fixed aperture used in CAMIRA for its richness estimate, while MCMF uses a scaling proportional to the expected size of r_{500} .

We summarize the comparison between CAMIRA and MCMF as follows. The photometric redshifts are highly consistent with each other when the same systems are matched. Outliers in redshift usually arise when the system found by MCMF has a low richness, which is where CAMIRA becomes incomplete. Judging from the last panel of Fig. 4, this is in the range of $20 < \lambda < 40$, depending on the cut in CAMIRA richness. The smaller footprint covered by HSC alone and the patchiness of the HSC data at the footprint visible in Fig. 1 can explain why

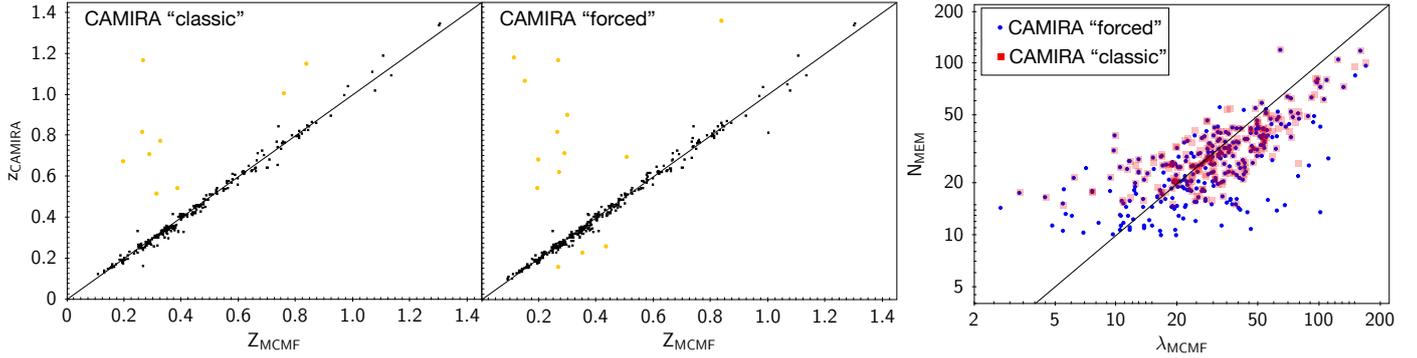


Fig. 4. Redshift and richness comparison between CAMIRA and MCMF. *Left:* photo- z comparison between default CAMIRA and MCMF and the same comparison using the forced CAMIRA run. Outliers are highlighted as larger colored points. *Right:* comparison of the richness estimate of CAMIRA (N_{MEM}) and MCMF λ for sources with consistent redshifts. Continuous lines show the one-to-one relation.

only 324 out of 542 candidates have a potential CAMIRA-based counterpart. We therefore used MCMF results for the baseline confirmation and redshifts and provide CAMIRA estimates only as additional information in the catalog.

4.2. General properties of the MCMF-confirmed cluster catalog

In total, we find 470 (446) clusters with $f_{\text{cont}} < 0.3$ (0.2), and $\sim 60\%$ of these systems have a spectroscopic redshift. The redshift distribution shown in Fig. 5 peaks at $z = 0.3$ and reaches out to $z = 1.3$. To provide some indication of the mass and redshift reach, we show in Fig. 6 the distribution of the sources in redshift versus approximate M_{500} , where the approximate M_{500} is that used in Sect. 3.1 to estimate cluster richness. Color-coded are matches to the SZ-based ACT-DR5 cluster catalog (Hilton et al. 2021) and matches to the X-ray-based ROSAT CODEX catalog (Finoguenov et al. 2020). This highlights the gain in terms of depth and number of clusters compared to the predecessor ROSAT and its sensitivity to high-redshift clusters compared to current SZ-based cluster surveys. Highlighted in yellow are eFEDS clusters found by matching with spectroscopy-based group catalogs (Sect. 5.1 and Table 1) that have failed confirmation by MCMF. Details for the matches between eFEDS and other catalogs are provided in Sect. 5.1. The distribution illustrates the shape of the X-ray selection for a sample following a flux limit and the high-sensitivity to low-mass systems at low redshift.

Finally, we show in Fig. 7 the distribution of candidates in the $\lambda - \text{approx.}M_{500}$ plane. Highlighted in green are confirmed systems, while unconfirmed systems are shown in blue. Although f_{cont} does not use this scaling relation to clean the candidate list, the cleaning obviously tends to exclude systems below the main relation. We further highlight in magenta sources that were not confirmed by MCMF but were found by cross-matching with group catalogs. The richness measurements for these systems were performed by fixing the redshift to the redshift of the corresponding group. A decent scaling between the two observables appears to exist down to the group mass regime.

Throughout the paper we define the cluster catalog by cutting the cluster candidate list at a certain f_{cont} (typically 0.2 or 0.3). We provide f_{cont} estimates for all candidates to allow users to adapt the level of catalog contamination and completeness to their needs. We do not recommend applying a higher threshold than 0.3 because the vast majority of $f_{\text{cont}} > 0.3$ systems are not clusters, and the redshifts provided are consequently likely

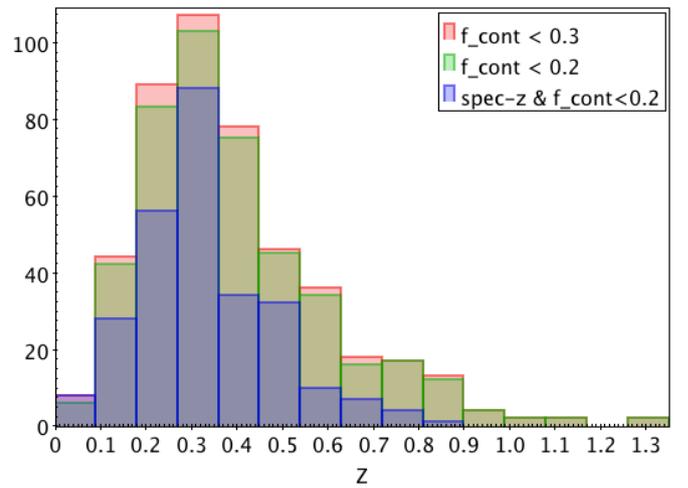


Fig. 5. Redshift distribution of the MCMF-confirmed eFEDS clusters for two cuts in f_{cont} and for the subset of clusters with $f_{\text{cont}} < 0.2$ and spectroscopic redshifts.

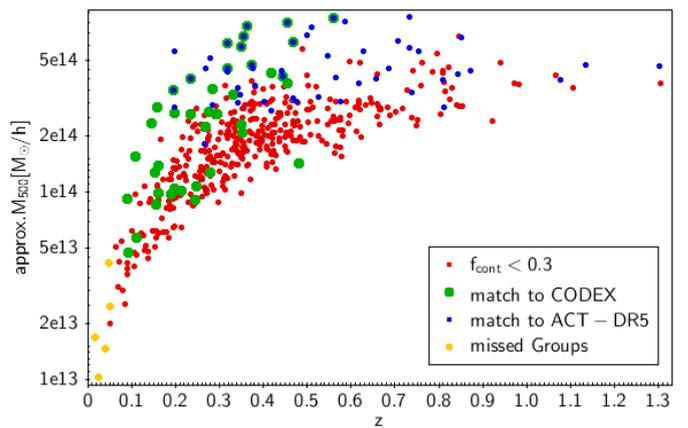


Fig. 6. Approximate M_{500} vs. redshift plot for $f_{\text{cont}} < 0.3$ clusters. Only 81 of the 470 systems show a match in the ICM-based CODEX and ACT-DR5 surveys. This illustrates the increase in the number of ICM selected clusters over the eFEDS field.

not associated with a true X-ray selected cluster. As outlined in Sect. 5.1.5, we manually added known missed systems to the sample by flagging them with f_{cont} values of -2 and -1 . With this, we expect the $f_{\text{cont}} < 0.3$ cluster catalog to include $>99\%$

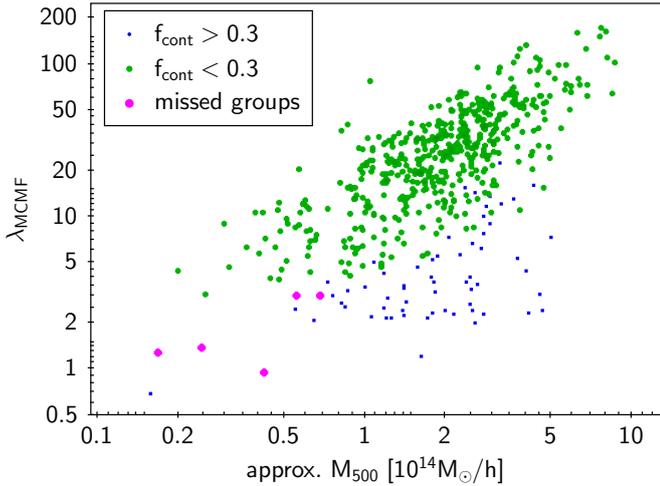


Fig. 7. Richness vs. approximate M_{500} for eFEDS sources. MCMF-confirmed sources ($f_{\text{cont}} < 0.3$) are shown in green, and $f_{\text{cont}} > 0.3$ systems are shown in blue. MCMF-unconfirmed (missed) groups found by catalog cross-matching are shown in magenta using a forced richness estimate at the group redshift.

of the true clusters in the candidate catalog with an estimated contamination level of 6% by noncluster sources.

4.3. Estimates of the optical dynamical state

The different estimators that depend on the shape of the galaxy distribution and its location relative to the X-ray center are well correlated. Two examples are shown in Fig. 8 for a complete set for all combinations plotted in Fig. C.1. With the lack of a simulation-based calibration, a simple way to select merging systems is selecting the most unrelaxed systems from a simple linear combination of these estimators. In Figs. 9 and 10 we show two of the most unrelaxed clusters with richness $\lambda > 50$.

In Fig. 9 we show the region around eFEDS J085620.7+014649 at $z = 0.269$, including three other detected extended sources. The main system shows clear indications of merging, with a complex morphology in X-ray surface brightness. The main cluster harbors at least two massive central galaxies and shows two peaks associated with two peaks in the X-ray surface brightness map. The second cluster merger shown in Fig. 10 is at $z = 0.73$. In X-rays, two distinct main clusters are both detected as extended. In the optical, both clusters appear to be connected by a bridge of galaxies, which is likely responsible for the classification of this system as an extreme merger from the estimators. The X-ray data are likely not deep enough to detect a possible connecting structure.

Another way to select interesting merger candidates is by examining the distance to the nearest neighbor. This can be done using both X-rays and optical data. In the right panel of Fig. 8, we show the distance to the nearest neighbor in optical versus nearest extended neighbor in X-rays in units of r_{500} . Sources aligned on the one-to-one line likely correspond to the same structures. A rather clean sample of cluster pairs can be obtained when low offsets in both X-ray and optical are required. We note here that cluster mergers can significantly impact the distribution of the ICM and its emission in X-rays. Having a nearby X-ray and optical neighbor may therefore cause such a selection to preferentially select cluster pairs in an early merging or a premerging phase. On the other hand, the X-ray source detection tool assumes circular models, and significant elongation or

complex morphology causes the code to identify multiple detections. In the right panel of Fig. 8, we highlight merger candidates in red if their nearest optical and X-ray neighbor is closer than $2.5 \times r_{500}$. This selection should include the majority of systems with overlapping virial radii. For convenience, these systems are listed in Table C.2.

In addition to providing group and cluster identification, it is well known that X-ray observations can be used to identify cluster mergers and to estimate their dynamical state. A study using eFEDS X-ray data is in preparation (Ghirardini et al. 2022), but was not yet ready for comparison with the optical estimates provided here. X-ray and optical estimators tend to probe different merger characteristics and are prone to different systematics. We therefore expect substantial scatter between different estimators of a dynamical state. On the other hand, the combination of information from X-ray and optical estimators can enable a selection of specific merger types. One example of this might be cluster pairs in pre- and post-collision state. Pairs can be selected with joint X-ray and optical information, where the X-ray data add the critical information about the dynamical state of the ICM, indicating a recent collision.

5. Catalog validation and performance of optical confirmation

In this section we investigate the performance of the optical confirmation by estimating the incompleteness due to optical cleaning and the final catalog purity. To do this, we performed a set of five steps or investigations, building upon the experience from the previous test. The first and most lengthy test is using cross matches to other catalogs. With this, we search for clusters and groups that are not confirmed by MCMF or show a discrepant redshift estimate. The second test compares the fraction of confirmed systems with those expected from simulated observations of the eFEDS field. The third check is an estimate of the incompleteness caused by the optical cleaning and a comparison with the missed systems found through cross-matching in the first test. For this, a scaling relation fit between the approximate M_{500} based on X-rays and the MCMF richness is performed. The fourth step consists of adding information from the tools dedicated to identifying point-like counterparts to X-ray candidates. This allows us to define cleaner or more contaminated subsamples of the cluster catalog. Finally, as a last step, we estimate the catalog contamination by using the scatter distribution around the derived scaling relation and a clean sample defined in the previous step. The results are then compared to those found in the previous steps.

5.1. Match to public catalogs

Cross-matching with public cluster catalogs allows us to assess the performance of the cluster confirmation and redshift assignment. The eFEDS field lies inside multiple surveys for which cluster and group catalogs were constructed. In particular, it overlaps the SDSS footprint, from which optical cluster catalogs were constructed using photometric as well as spectroscopic data. Furthermore, it overlaps the cluster catalogs of ACT-DR5 (Hilton et al. 2021) and is naturally covered by all-sky cluster surveys such as *Planck* PSZ2 (Planck Collaboration XXVII 2016) and ROSAT (Piffaretti et al. 2011).

Because so many cluster catalogs are available, we restrict ourselves to the largest or most commonly used catalogs and to a subset of possible tests. The focus of the tests is on clusters that

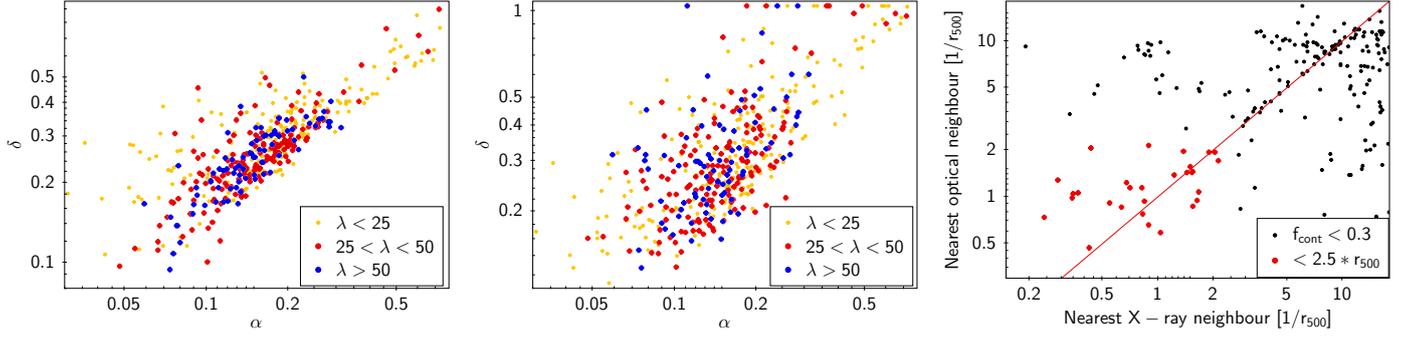


Fig. 8. Dynamical state estimators. *Left and middle panels:* comparison between the three estimators defined in Wen & Han (2013). Three different richness bins are color-coded. *Right:* distance to the next neighbor found in the galaxy density map vs. distance to the next neighbor in the X-ray cluster catalog. The one-to-one relation is indicated as a red line. Clusters showing a near neighbor within $\sim 2.5 * r_{500}$ in both X-ray and optical are likely cluster pairs that potentially are merging.

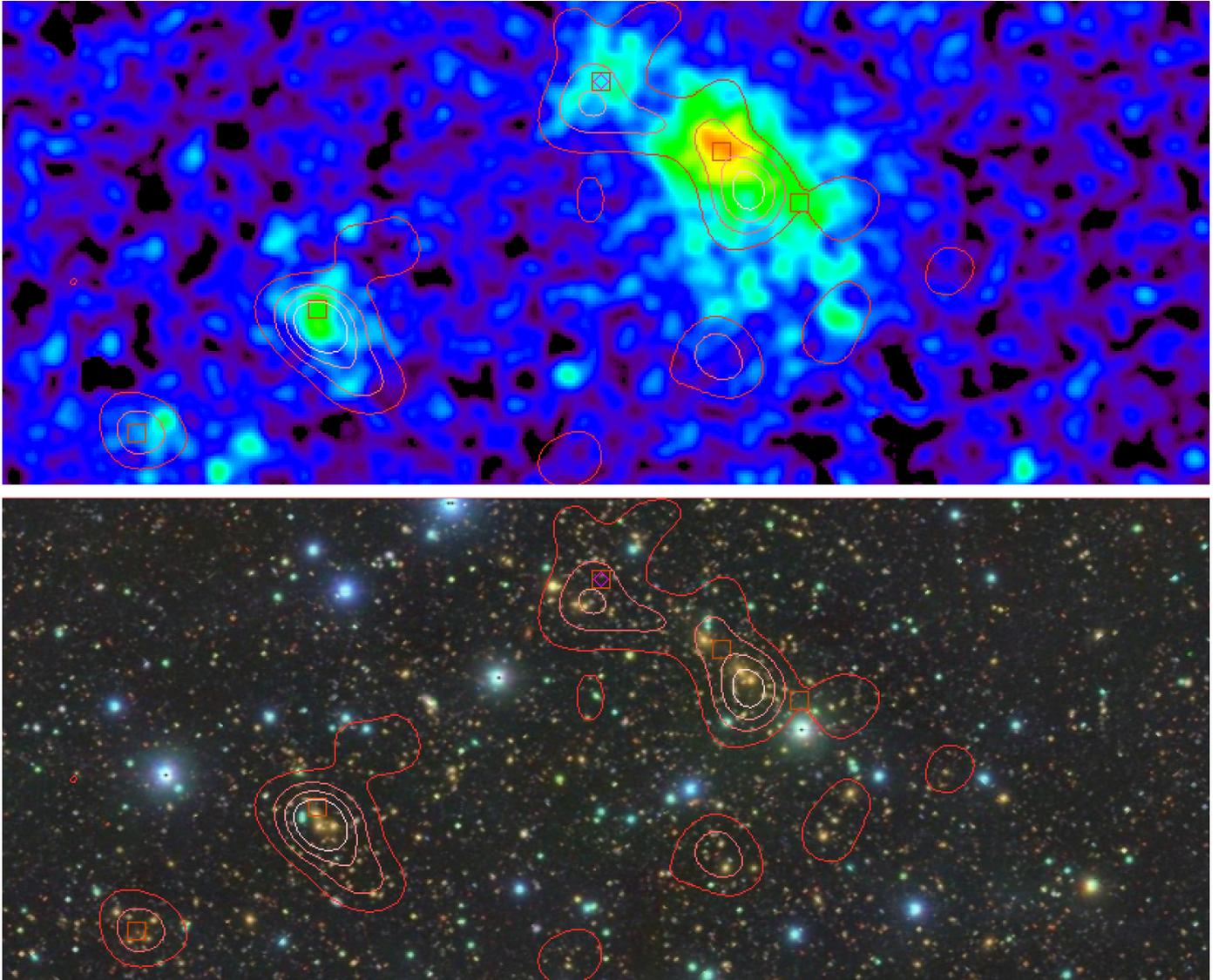


Fig. 9. Region around eFEDS J092220.4+034806, the first of the three most unrelaxed systems with $\lambda > 50$ according to optical dynamical state estimators. *Top:* smoothed 0.5–2.0 keV X-ray surface brightness map of a 7×2.8 Mpc region around the merging system. Contours show the redshift-based galaxy density at the cluster redshift. Boxes show extent-selected sources at the system redshift. *Bottom:* HSC *grz*-color composite image of the same region.

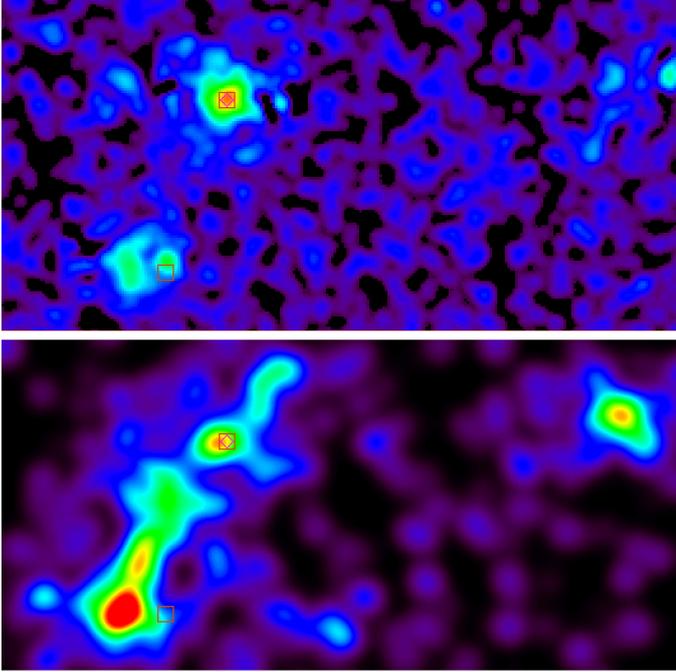


Fig. 10. Region around eFEDS J085620.7+014649, the second of the three most unrelaxed system with $\lambda > 50$ according to optical dynamical state estimators. *Top:* smoothed 0.5–2.0 keV X-ray surface brightness map showing two main clumps separated by 5 arcmin (2 Mpc). Boxes show extent-selected sources at the system redshift. *Bottom:* red-sequence galaxy density map at $z = 0.73$ over the same region. The galaxy density map suggests a connection between the two clusters. Additionally, a third cluster is clearly visible about 3 Mpc away from the main pair.

MCMF failed to confirm or where redshifts are misassigned. We do not aim to test the X-ray selection or the selection function of the other surveys with which we matched our clusters and groups.

We performed two standardized tests of the data: redshift consistency, and confirmation fraction of matched systems. Systems with inconsistent redshifts or high f_{cont} were confirmed individually. The summary is provided in Sect. 5.1.5.

5.1.1. X-ray selected clusters

The ROSAT all-sky survey (RASS; Truemper 1982; Voges et al. 1999) can be seen as a predecessor of the eROSITA all-sky survey (eRASS), which provides shallower X-ray observation over the whole sky with poorer angular resolution. While early work performed cuts to start with rather clean X-ray catalogs that need confirmation, more recent works (Klein et al. 2019; Finoguenov et al. 2020) make use of large optical photometric surveys to systematically identify clusters within all RASS detections. This approach results in cluster catalogs that are about ten times higher in source density than previous RASS-based catalogs.

We matched the eFEDS extended sources to two cluster catalogs: to the MCXC catalog (Piffaretti et al. 2011), and to the CODEX catalog (Finoguenov et al. 2020). We used a maximum distance between X-ray positions of two arcminutes.

For MCXC, we find only one source in the eFEDS footprint, RXC J0920.0+0102 at a redshift of $z = 0.017$, which is matched to the eFEDS source eFEDS J092002.1+010219. This cluster is not confirmed by MCMF, although a MCMF measurement on a substructure 130 arcsec from the cluster is close to being detected

($f_{\text{cont}} = 0.35$, $z = 0.033$). The very low redshift and estimated mass of $M_{500} = 0.36 \times 10^{14} h^{-1} M_{\odot}$ make it hard to confirm this cluster with MCMF. The deep-imaging data used in MCMF are not optimized for very low redshift sources with their large angular size. The size of the BCG shows a diameter of 5 arcmin in HSC, several hundred times larger than a typical galaxy in these surveys, and dozens of background galaxies shine through the BCG.

In the case of the CODEX catalog, we find 43 positional matches. RXC J0920.0+0102 is not part of this catalog. All positional matches have $f_{\text{cont}} < 0.02$ and are therefore clearly confirmed by MCMF. We find three matches with a redshift offset of more than 5%. In two cases the optical system found in CODEX shows a large offset from the eFEDS X-ray source. The MCMF redshifts of both eFEDS sources are greater than 0.6 and therefore beyond the redshift reach of the SDSS data used to construct CODEX. The remaining cluster eFEDS J091509.5+051521 belongs to an eFEDS source with two clusters along the line of sight with f_{cont} of 0.012 and 0.017, which we show in Fig. 11. The redshift of the second-ranked cluster is consistent with that of CODEX. The difference in f_{cont} is too small for a clear single redshift assignment. Using the richness as a mass proxy to predict the number of X-ray photons emitted by each cluster yields similar numbers for both clusters. It is therefore likely that both clusters significantly contribute to the X-ray detection.

Given the RASS-based CODEX catalog and the great improvement regarding depth and angular resolution of eROSITA over ROSAT, we also determined the matched fraction of CODEX clusters over eFEDS. To do this, we repeated the matching with a more generous cut on the offset of 3 arcmin and limited the CODEX catalog to sources well within the eFEDS footprint to avoid border effects. We furthermore allowed multiple sources to match one source so that the results are less affected by source splitting in either catalog. We find that only 45% of the CODEX clusters match an extended eFEDS source. With the cleaning flag in CODEX, which imposes a redshift-dependent richness cut similar to f_{cont} , the match fraction increases to 85% and the number of matched sources decreases by 30%. The unmatched CODEX sources in the clean subsample are usually caused by mismatches between the optical source and the X-ray source. As an example, the X-ray source given in CODEX is obviously a bright point source found in eFEDS. At the position of the associated optical cluster, however, X-ray emission is indeed found, but this source is not the same X-ray source as is listed in CODEX. Expanding the examination to the nonclean sample of nonmatched sources reveals a further increased fraction of mismatches mostly with bright point sources and an increased number of sources for which no X-ray source is found at the X-ray or the optical position given in CODEX.

If the eFEDS region is representative of the full CODEX catalog, then the CODEX catalog contains more than 50% of the sources in which the X-ray flux is not associated with a real X-ray cluster. When the cleaning flag in CODEX is applied, however, the catalog becomes much cleaner, with a purity of 85% or higher, including the fact that most of the nonmatched clean sources indeed show X-ray emission at the optical position.

In summary, the match with CODEX raises our awareness of multiple structures along the line of sight and the importance of the redshift limits of the optical survey. The match with MCXC reveals a missed group at low redshift, highlighting the difficulty of confirming very nearby low-mass systems with photometric surveys.

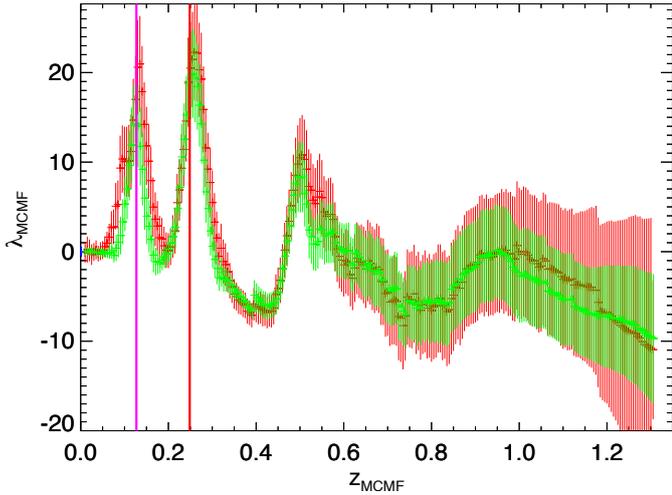


Fig. 11. MCMF richness vs. redshift for eFEDS J091509.5+051521, one of the cases with a photo- z mismatch to the CODEX cluster catalog. MCMF finds two counterparts that are highlighted as vertical lines with f_{cont} of 0.012 (red) and 0.017 (magenta). The redshift given in CODEX corresponds to the magenta line. The richness from the LS $grzW1$ -band run to predict X-ray count rates yields similar values for both peaks. Red points show LS grz , and green shows the LS $grzW1$ bands. This cluster is not in the HSC-SSP footprint.

5.1.2. SZ-selected surveys

The eFEDS survey overlaps the *Planck* and the ACT SZ surveys. The SZ effect is redshift independent, and depending on the frequency, it can be observed as a negative signal imprinted on the cosmic microwave background. The selection functions of both surveys (*Planck*, ACT) are very different because detection also depends on angular resolution and available frequency range. This results in a strongly redshift-dependent selection in *Planck* similar to that seen in flux-limited X-ray surveys, while ACT shows a flat selection with redshift that becomes even more sensitive at higher redshift as beam size and cluster size start to match.

We matched the eFEDS catalog to the *Planck* PSZ2 catalog (Planck Collaboration XXVII 2016) within 3 arcmin and found ten matches. All matches show consistent redshifts and $f_{\text{cont}} = 0$.

Matching the eFEDS catalog to the ACT-DR5 catalog, we find 53 matches. All matches have $f_{\text{cont}} < 0.07$ and therefore are MCMF-confirmed systems. This includes three eFEDS clusters with $z > 1.0$, including the highest-redshift eFEDS cluster with $z = 1.3$.

We initially find three photo- z mismatches. eFEDS J084441.3+021702 (ACT-CL J0844.6+0216) was assigned a spectroscopic redshift in ACT-DR5 that was assigned to a foreground galaxy at $z = 0.56$. The MCMF-identified cluster has a BCG with spectroscopic redshift $z = 0.6515$. Here we expect the MCMF redshift to be the correct one.

The second mismatch, eFEDS J083120.5+030949 (ACT-CL J0831.3+0310), appears in the X-ray map as two structures separated by 50 arcsec, but this is detected as one cluster with its center in the middle and is shown in Fig. 12. MCMF assigns $f_{\text{cont}} = 0$ to both clusters, and their richness is 50.1 for the $z = 0.852$ cluster and 47.9 for the $z = 0.569$ cluster. As both have the same f_{cont} , the ranking of the clusters is somewhat arbitrary. The SZ center and redshift ($z = 0.566$) fits the cluster at lower redshift well. The similar richness of both peaks further suggests

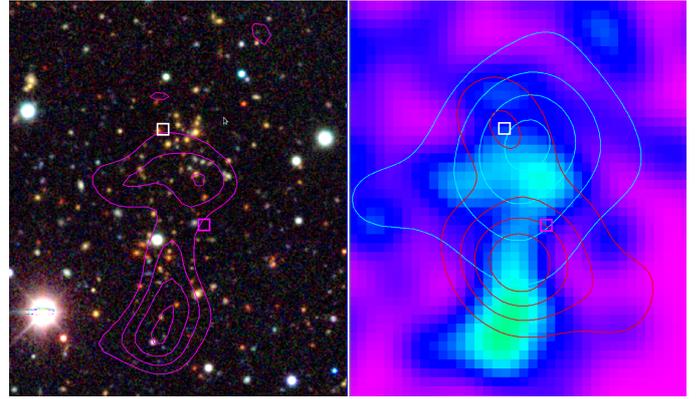


Fig. 12. 2.5×3 arcmin region around eFEDS J083120.5+030949 (ACT-CL J0831.3+0310). *Left:* LS grz -band color image. Magenta contours are drawn from the X-ray surface brightness map, shown in the right panel. The magenta box shows the eFEDS position, and the white box shows the SZ position from ACT. *Right:* X-ray map, showing X-ray and SZ position with density contours for passive galaxies at $z = 0.569$ (cyan) and $z = 0.852$ (red).

that the majority of the X-ray photons are emitted from the cluster with the lower redshift. We therefore change the order of the two peaks, assigning the lower-redshift cluster as the primary counterpart to the X-ray selected clusters.

The eFEDS cluster eFEDS J091646.9+015531, with the positional match ACT-CL J0916.7+0155, has two optical counterparts with similarly low f_{cont} of 0.027 and 0.036. The richness versus redshift plot is shown in Fig. 13. The second-ranked counterpart with $z = 1.17$ fits the value listed in ACT of $z = 1.15$ well. Given the different cluster sensitivity with redshift of X-rays and SZ, it is possible that despite the very close positional matches, the clusters in both surveys see are in fact different. This scenario appears to apply here. The X-ray and richness fit the lower-redshift cluster well. The X-ray based approximate M_{500} places all flux at $z = 1.17$, however, which suggests a three times higher mass than is found by ACT. The richness for the $z = 1.17$ found by MCMF is consistent with the ACT-based mass estimates. The richness of the low-redshift cluster is about half of that at high redshift. When we assume that the richness ratio is similar to the mass ratio of the clusters, the ACT signal would indeed be dominated by the high-redshift cluster. We conclude that although the redshifts of the ACT cluster and the eFEDS cluster disagree, the redshift assignments appear to be correct for both cases.

5.1.3. Match to optical cluster catalogs

In addition to its direct access to the most recent HSC-based CAMIRA catalog, the eFEDS field overlaps the SDSS and KiDS surveys that were used to construct cluster catalogs. In the case of the SDSS, the number of public cluster catalogs is indeed quite large. We therefore restrict our analysis to only three catalogs: redMaPPer (Rykoff et al. 2016) on SDSS DR8, AMF (Banerjee et al. 2018) on SDSS DR9, and WHL (Wen & Han 2015) on SDSS DR12.

The match to redMaPPer was performed within a 2 arcminute radius. We find 163 matches. All systems have $f_{\text{cont}} < 0.3$, and all except one show $f_{\text{cont}} < 0.08$. The match with $f_{\text{cont}} = 0.29$ is about 90 arcsec from the X-ray position at a location without X-ray emission in the map. We therefore classify this match as a chance match given the adopted search region.

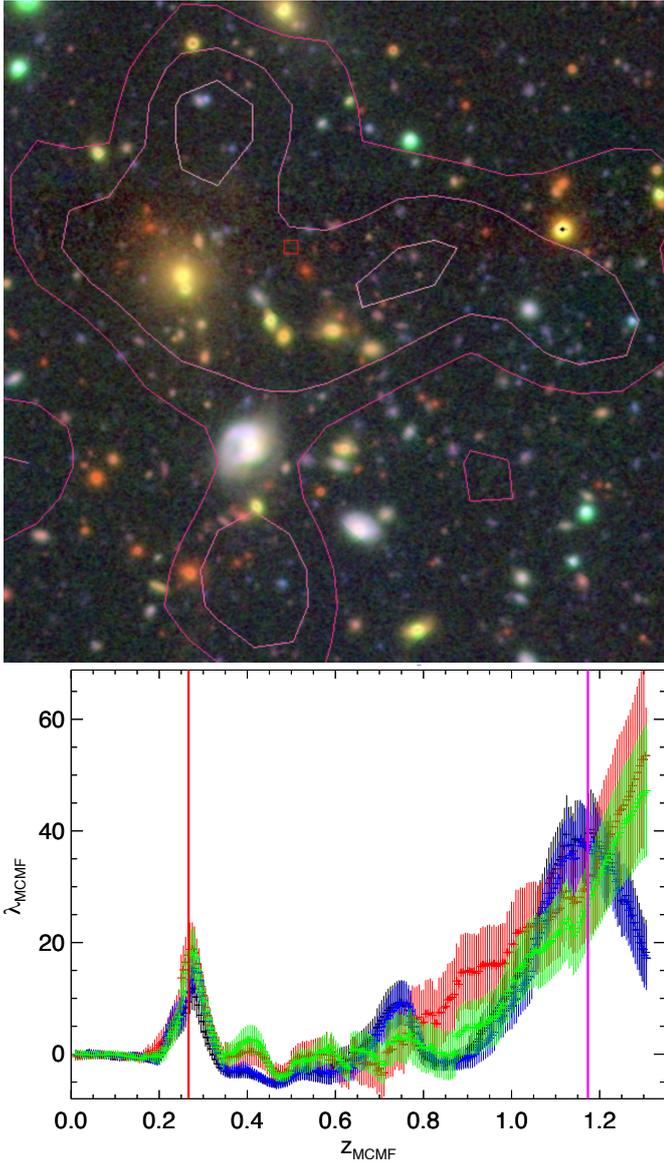


Fig. 13. X-ray source eFEDS J091646.9+015531. *Top:* HSC *grz*-color image of the central 2.5×2.5 arcmin region with X-ray contours overlaid. *Bottom:* MCMF richness vs. redshift plot, one of three cases of a photo- z mismatch with positional matches from the ACT-DR5 cluster catalog. MCMF finds two counterparts that are highlighted with vertical lines. The richer system at $z = 1.17$ is consistent with the ACT cluster. The different sensitivity to redshift for X-rays suggests that the low redshift is the better counterpart to the eFEDS source. Colors indicate different MCMF runs: HSC S19A (black), HSC S20A (blue), LS *grz* (red), and LS *grzw1* (green) bands.

We find only one outlier in redshift, eFEDS J091509.5+051521, which was already discussed in Sect. 5.1.1 as two systems along the line of sight. Because redMaPPer on SDSS was used to create the CODEX catalog, this mismatch is expected.

The match to AMF was performed using the same maximum offset as for redMaPPer. We find 129 sources. All show $f_{\text{cont}} < 0.3$, and only one shows $f_{\text{cont}} > 0.2$. The eFEDS cluster with the highest f_{cont} eFEDS J084004.8+013751 shows a large X-ray to BCG offset of 102 arcsec. Judging from the X-ray surface brightness map, the X-ray source associated with the BCG and AMF counterpart appears to be separated from the eFEDS-detected X-ray source and is not associated with a source in

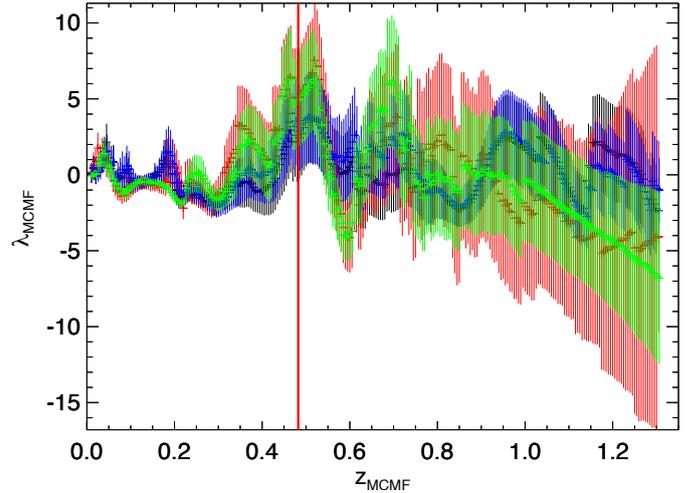


Fig. 14. Richness versus redshift plot of eFEDS J084430.8+021736, a galaxy group at $z = 0.05$ that is not confirmed by MCMF and found by cross-matching with the WHL catalog. The color-coding is the same as in previous richness vs. redshift plots.

the extent-selected sample. We find seven matches with photo- z offsets greater than 5%. For four sources, AMF matches the second-ranked counterpart. Two sources show AMF redshifts between the second- and first-ranked MCMF counterpart. The small redshift differences between the first and second MCMF peaks of $\Delta z < 0.2$ may have caused blending issues in the AMF redshift estimation. The last cluster with a photo- z mismatch has an MCMF photo- z and a spec- z of $z = 0.61$ and therefore lies beyond the reach of AMF, which assigned $z = 0.51$. In summary, all AMF matches are associated with MCMF-confirmed systems, and the redshift mismatches typically stem from a different ranking of multiple possible counterparts or a photo- z limitation of the AMF catalog.

Repeating the same exercise with the WHL catalog, we find 244 matches, 243 of which are matched to $f_{\text{cont}} < 0.3$ sources and 239 show $f_{\text{cont}} < 0.2$. Of these five $f_{\text{cont}} > 0.2$ systems, three have consistent redshift and richness with the MCMF estimates. One outlier is eFEDS J084004.8+013751, which was already discussed in the AMF cross match. The last source, the only system with $f_{\text{cont}} > 0.3$, is eFEDS J084430.8+021736, which is a $z = 0.05$ group with an approximate mass of $M_{500} = 2.45 \times 10^{13} h^{-1} M_{\odot}$. The λ versus redshift plot in Fig. 14 shows a peak of richness 1–2 at redshift 0.05. Given the approximate X-ray based mass, the indicated richness is close to the expectation and illustrates the difficulty of identifying eFEDS systems in the low-redshift low-mass regime with optical photometry. Additionally, the rich eFEDS cluster eFEDS J084441.3+021701 lies just 2.5 arcmin away from this group, and the assigned MCMF redshift of eFEDS J084430.8+021736 corresponds to the residual signal of that cluster. We find seven redshift outliers, including the aforementioned group at $z = 0.05$ and the double system eFEDS J091509.5+051521. In three of the remaining cases, we find a second-best counterpart with $f_{\text{cont}} < 0.1$ consistent with WHL. The remaining two matches are obvious chance matches, where the WHL system is more than 90 arcsec away without associated X-ray emission.

5.1.4. Group surveys

As already indicated by the missed confirmation of RXCJ0920.0+0102 discussed above, it becomes hard to

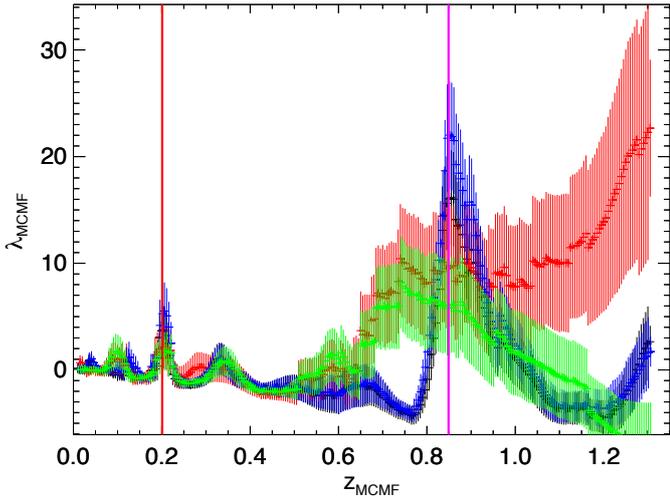


Fig. 15. eFEDS J090806.4+032613. The color-coding is the same as in the previous richness vs. redshift plots.

automatically confirm very nearby X-ray selected systems using the photometric data from HSC or DECALS. The two main reasons for this are issues with the photometry of these nearby galaxies and the low mass of these systems. Figures 6 and 7 show that the confirmed X-ray clusters reach down to low-mass groups for which it is hard to distinguish between projected and bound galaxy over density using photometry alone. Group and cluster catalogs based on spectroscopy allow probing this regime to evaluate and recover missed systems. We restrict this comparison to two catalogs based on different surveys, one using SDSS data (Tempel et al. 2014), and the other using redshifts from 2MASS-selected galaxies.

With the available spectroscopic redshifts from SDSS, various group catalogs were constructed. We selected the group catalog based on SDSS DR10 from Tempel et al. (2014). We find 35 matches within a 120 arcsec matching radius. Of these, 33(32) show $f_{\text{cont}} < 0.3(0.2)$. Of 33(32) sources with low f_{cont} , 25 have consistent redshifts, while all sources with $f_{\text{cont}} > 0.3$ have inconsistent redshifts. Of the 10 systems with inconsistent redshifts, we recovered RXC J0920.0+0102, the $z = 0.017$ group discussed in Sect. 5.1.1, as well as the multiple system eFEDS J091509.5+051521 that we discussed in the same section. Of the remaining clusters, only eFEDS J092821.1+041941 has an incorrectly assigned redshift. In this case, MCMF assigns the redshift of a neighboring more massive cluster (eFEDS J092821.2+042149) to the X-ray source, while the actual best match is ranked second. In addition to RXC J0920.0+0102, the only unconfirmed system with a match to the SDSS group catalog is eFEDS J090806.4+032613. The optical investigation of this source does not provide a clear answer. The mass according to the SDSS catalog of the matched group is $1.8 \times 10^{13} h^{-1} M_{\odot}$, which is at the lower end of the limits that can be detected in eFEDS. The three associated member galaxies are widely spread. 1.2 arcmin (140 kpc) is the distance to the nearest member to the X-ray center. Furthermore, structures are found in MCMF at $z = 0.2$ and $z = 0.84$, which both contain at least one galaxy with a spectroscopic redshift (see Fig. 15). We therefore categorize the system as an unclear case.

Of the available 2MASS-based group catalogs, we used the catalog provided in Tully (2015) because it provides the largest number of matches within 120 arcsec with the eFEDS candidates. We find 7 positional matches, 5 of which show offsets

of less than 10 arcsec, while the other two show offsets greater than 100 arcsec. One of the two outliers is again associated with RXC J0920.0+0102. The other match with a large offset is likely a chance association because no significant X-ray emission appears to be associated with the group center and the offset corresponds to more than ten times the measured extent of that source. All matches show $f_{\text{cont}} > 0.3$ and are therefore not considered as MCMF confirmed, while all 5 matches with small positional offset can be considered as X-ray detected groups. The missed systems show redshifts up to $z = 0.05$ and reach X-ray luminosity based masses of up to $M_{500} \approx 6.0 \times 10^{13} h^{-1} M_{\odot}$.

5.1.5. Summary of the cross-matching results

The cross match with X-ray catalog resulted in the identification of one missing system and one system with two counterparts. All clusters matched to SZ selected systems are confirmed, but the redshift assignment of one cluster, eFEDS J083120.5+030949, is changed to the initially second-ranked cluster that shows similarly low f_{cont} . From the match with optical photometric cluster catalogs, we identify one system, eFEDS J084430.8+021736, that was missed by MCMF. The match to spectroscopic group and cluster catalogs yielded one redshift reassignment, eFEDS J092821.1+041941. Furthermore, it revealed five additionally missed groups that match the 2MASS-based group catalog by Tully (2015). These seven missed systems are listed in Table 1 and were manually added to the cluster catalog. We assigned dummy f_{cont} values of -2 for group matches where visual investigation of the optical and X-ray images yielded X-rays from extended gas or from individual galaxies. A value of -1 was assigned for group matches that showed a clear indication of extended X-ray emission. In Fig. 16 we show the distribution of confirmed and unconfirmed eFEDS cluster candidates in extent likelihood (EXT_LIKE) versus detection likelihood (DET_LIKE), two of three key source selection variables. The third variable, source extent, modulates the width of the observed distribution. The systems found by the matching exercise that are listed in Table 1 are shown as magenta squares. In addition to the obvious scaling relation between EXT_LIKE and DET_LIKE, it becomes obvious that unconfirmed systems cluster at low EXT_LIKE and DET_LIKE. Three of the missed groups lie in the very high regime in EXT_LIKE and DET_LIKE, where contamination by nonclusters is highly unlikely. All MCMF-unconfirmed systems that are not confirmed by cross-matching with group catalogs lie within $\text{EXT_LIKE} < 20$ and $\text{DET_LIKE} < 80$.

5.2. Confirmation fraction and incompleteness

One way to evaluate the performance of MCMF cleaning is to compare the observed fraction of confirmed clusters to the expected confirmation fraction or purity of the X-ray catalog based on simulations. In the case of eFEDS, dedicated simulations were performed that included details of the eFEDS footprint, background, and exposure times. A detailed description of the simulations and their results are shown in Brunner et al. (2022); Liu et al. (2022b). One result is that the purity of the catalog is a strong function of extent likelihood (EXT_LIKE), which allows us not only to compare the overall confirmation fractions of simulation and observation, but also to verify whether its behavior as a function of extent likelihood is as expected.

Table 1. MCMF unconfirmed systems ($f_{\text{cont}} > 0.3$) that have a counterpart in one of the matched catalogs.

NAME	RA	Dec	DET_LIKE	EXT_LIKE	z	M_{500}	f_{cont}	λ
eFEDS J092002.1+010220	140.0090	1.0388	1049.64	179.60	0.017	0.169	-1.0	1.30
eFEDS J093744.1+024536	144.4340	2.7600	66.58	11.84	0.024	0.103	-2.0	0.02
eFEDS J090811.6-014811	137.0486	-1.8032	72.52	13.37	0.04	0.146	-2.0	0.00
eFEDS J084034.5+023638	130.1440	2.6107	661.78	114.64	0.049	0.42	-1.0	0.95
eFEDS J084430.8+021736	131.1284	2.2935	79.31	27.49	0.0504	0.245	-2.0	1.40
eFEDS J084531.6+022831	131.3818	2.4753	275.71	98.86	0.0765	0.68	-1.0	3.00
eFEDS J093141.2-004717	142.9219	-0.7883	30.46	13.71	0.093	0.556	-1.0	3.00

Notes. The f_{cont} entries are set to -1 (good) and -2 (less good) counterparts. The approximate X-ray based mass estimate is given in units of $10^{14} M_{\odot} h^{-1}$. The richness λ is obtained at the given redshift.

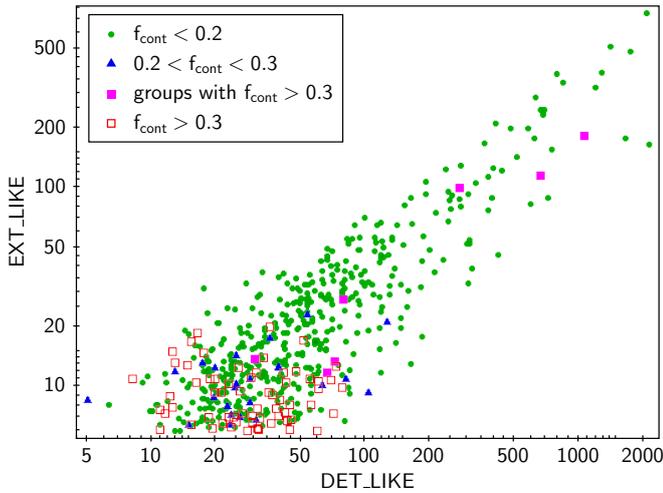


Fig. 16. Distribution of X-ray cluster candidates in X-ray extent likelihood vs. detection likelihood, two of three key selection parameters. Clusters confirmed by MCMF are shown in green and blue. Unconfirmed clusters are shown in red. Systems not confirmed by MCMF but via cross-matching with group catalogs are shown as magenta squares. Unconfirmed systems lie at low likelihood values where contaminants are expected.

In Fig. 17 we show the results of this comparison. The dashed magenta line shows the fraction of extent-selected X-ray detections associated with clusters or cluster wings over the number of all extent-selected X-ray detections in the simulation. Sources classified as cluster wings are associated with a cluster, but are not the primary detection. Because we did not filter for such cases in the optical confirmation, we included them in the overall fraction of confirmed systems. The blue line shows the fraction of $f_{\text{cont}} < 0.3$ systems corrected for the residual contamination allowed,

$$f_{\text{corrected,EXT_LIKE}} = \frac{N_{\text{EXT_LIKE}}(f_{\text{cont}} < c)}{N_{\text{EXT_LIKE}}} (1 - c c_{\text{Sim}}), \quad (7)$$

where $N_{\text{EXT_LIKE}}$ is the number of candidates in a given bin in EXT_LIKE, $N_{\text{EXT_LIKE}}(f_{\text{cont}} < c)$ is the subset with f_{cont} below a threshold c (here 0.3), and c_{Sim} is the fraction of nonclusters in the simulation. The green line shows the same, but for $f_{\text{cont}} < 0.2$ instead.

The difference between the lines derived using $f_{\text{cont}} < 0.3$ and $f_{\text{cont}} < 0.2$ indicates excess incompleteness caused by the stricter cut. The difference between the two cuts means that

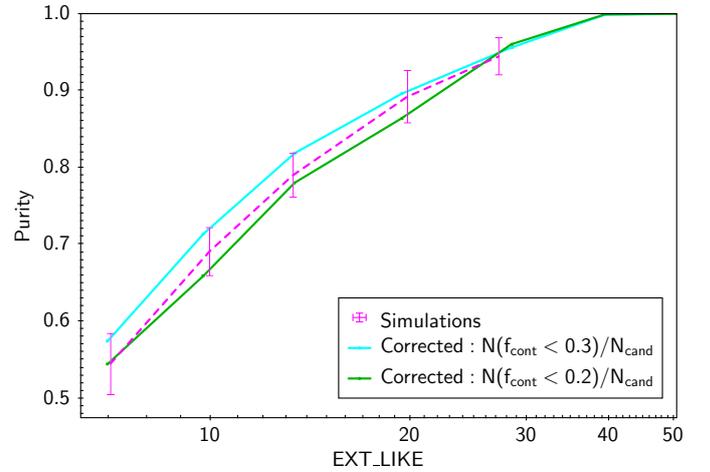


Fig. 17. Expected purity and recovered confirmation fraction as a function of extent likelihood. The expected fraction of real clusters from dedicated eFEDS simulations is shown in magenta. The error bars indicate the standard deviation derived from 18 realizations of the simulations. Blue shows the fraction of $f_{\text{cont}} < 0.3$ systems corrected for the expected number of residual contaminants. Green is similar, but for $f_{\text{cont}} < 0.2$. The difference between the green and blue line is an indication of the incompleteness caused by the stricter f_{cont} selection.

14 (3%) of the real sources are missed with the stricter cut, which is within the scatter range seen within the simulations.

Using the curves in Fig. 17 and the number of sources given EXT_LIKE, we can estimate the contamination fraction of the original cluster candidate list under the assumption that no incompleteness is caused by the cleaning. Consequently, using the curve including the identified groups that are missed by MCMF, we expect 442 real sources; given the 542 candidates, this results in a contamination fraction in the original candidate list of 17%.

5.3. Estimate of the excess incompleteness induced by optical cleaning

Any cluster survey that needs auxiliary information such as redshifts or cleaning of contaminants likely has an additional incompleteness that stems from the process of obtaining the needed information. This is especially true if the initial cluster survey challenges the redshift or mass reach of the follow-up survey. In this particular case, eFEDS challenges the optical confirmation with LS and HSC in the low-mass and low-redshift regime. This is already reflected in the redshifts and masses of

the systems that we find were missed by MCMF (discussed in Sect. 5.1.5). The MCMF confirmation or cleaning of the X-ray candidate catalog is based on a redshift-dependent richness cut that corresponds to a cut in f_{cont} . Because this is a systematic confirmation method, its impact can be modeled using scaling relations that connect detection observable and richness.

The proper way to measure and account for the impact of the optical cleaning on the sample is to include the richness measure in the overall analysis and trace its scaling with mass and selection observable simultaneously. For an example using an MCMF-based catalog, see Grandis et al. (2020). We performed a simplified estimation of the incompleteness induced by the MCMF cleaning by only using observed quantities.

As a first step, we measured the $\lambda_{\text{MCMF}} - M_{500}$ scaling relation and its scatter. We note here again that M_{500} is the approximate mass estimate based on a simple model, and it is therefore close to the count rate and in turn to the detection probability, just taking out the most pronounced dependences such as on redshift and Galactic nH column density. We assumed that the distribution of richnesses λ at a given mass M_{500} and redshift z is given by a log-normal distribution,

$$P(\ln \lambda | M_{500}, z) = \mathcal{N}(\ln \lambda; \langle \ln \lambda \rangle(M_{500}, z), \sigma^2(M_{500}, z)), \quad (8)$$

with mean

$$\langle \ln \lambda \rangle(M_{500}, z) = \ln \lambda_0 + \alpha_0 + \alpha_M \ln \left(\frac{M}{M_0} \right) + \alpha_z \ln \left(\frac{1+z}{1+z_0} \right), \quad (9)$$

and variance

$$\sigma^2(M_{500}, z) = \exp \left(\ln \zeta(z) - \langle \ln \lambda \rangle \right) + \exp \left(s + s_M \ln \left(\frac{M}{M_0} \right) \right), \quad (10)$$

where $(\alpha_0, \alpha_M, \alpha_z, s, s_M)$ are the free parameters to be constrained by the likelihood analysis, $(\lambda_0 = 23, M_0 = 3e14 M_\odot h^{-1}, z_0 = 0.35)$ are the pivots of the scaling relation and are chosen close to the medians in richness, mass, and redshift, respectively. $\zeta(z)$ is a redshift-dependent factor that relates the number of actually measured galaxies to the reported richness (see Eq. (7) in Klein et al. 2019). It modulates the first term of the variance, which captures the Poisson noise in measuring the richness. The second term of the variance models the intrinsic variance of the cluster population. We find best-fit parameters of $\alpha_0 = 0.192 \pm 0.025$, $\alpha_M = 1.14 \pm 0.06$, $\alpha_z = -1.25 \pm 0.26$, $s = -1.69 \pm 0.10$, and $s_M = -0.12 \pm 0.17$. More details about the fitting, including the impact of optical cleaning, can be found in Appendix A.

Based on the scaling relation, we can calculate for each confirmed cluster i the probability that it is not confirmed as

$$P_i = \int_{-\text{inf}}^{\ln \lambda_{\text{cut}}(z_i)} d \ln \lambda P(\ln \lambda | M_{500,i}, z_i), \quad (11)$$

where $\ln \lambda_{\text{cut}}(z_i)$ is the richness cut for a given f_{cont} threshold and cluster redshift z_i . With this information, we can derive a first estimate of the number of missed systems by summing over all clusters as

$$N_{\text{missed}} = \sum_i P_i. \quad (12)$$

Using the posteriors from the scaling relation and taking the 6% (4%) residual contamination for the f_{cont} 0.3 (0.2) into account, we end up with $10.4^{+2.6}_{-2.6}$ ($25.8^{+6.0}_{-4.7}$) missed systems.

Although these numbers fit the observed number of unconfirmed systems as well as the expected loss of systems when going from $f_{\text{cont}} < 0.3$ to $f_{\text{cont}} < 0.2$ (14 systems) quite well, they do not make use of the observed count rates and the number of unconfirmed candidates in the candidate list. The numbers further rely on a reasonably sampling of the redshift range of true clusters, which might be violated because we do not confirm any cluster below $z < 0.05$, while the majority of missed systems are at that redshift. We therefore employed a second more sophisticated approach that uses all candidates, the observed purity given extension likelihood, and the redshift distribution. We modified Eq. (11) by replacing the cluster redshift z_i by the normalized smoothed redshift distribution $P_z(s)$ of confirmed clusters plus the missed systems that were added after cross comparison. The assumption here is that this distribution is close to the distribution of the true clusters in the X-ray catalog. To account for catalog contamination, we further multiplied this probability by P_{true} , the probability of a candidate to be true given the extent likelihood, as shown in the previous section. With this, the probability of a system to be missed is

$$P_j = P_{\text{true},j} \int_0^{\text{inf}} dz \int_{-\text{inf}}^{\ln \lambda_{\text{cut}}(z)} d \ln \lambda P(\ln \lambda | M_{500,j}, z) P_z(z). \quad (13)$$

Summing now over all X-ray candidates yields $8.2^{+2.2}_{-1.5}$ and $22.5^{+4.2}_{-3.0}$ for $f_{\text{cont}} < 0.3$ and 0.2 and a difference between the two of $14.2^{+2.2}_{-1.3}$. This also fits the number of missed systems well, the difference in number between when the f_{cont} cuts are shifted and the simpler estimate before.

As a side result, we can further estimate the redshift and mass distribution of the missed systems by simply omitting the integral over redshift and summing over all candidates. The redshift distribution is shown in Fig. 18 and peaks at very low redshifts, with a flat tail over the remaining redshift range. The missing systems that were later found by cross-matching are consistent with the peak of the redshift distribution, and the same is true for the mass distribution shown in Fig. 19.

In summary, this approach of estimating the incompleteness induced by the optical confirmation process provides consistent results between findings from individual catalog matches and the difference of confirmed systems for different f_{cont} cuts. Furthermore, the predicted redshift distribution of the missed systems is consistent with that of the missed systems that are found by cross-matching to other catalogs. We find that the optical cleaning induces an incompleteness of 2% (5%) for $f_{\text{cont}} < 0.3$ (0.2) cuts. Based on this, adding the identified MCMF-unconfirmed but spectroscopically confirmed groups to the list of confirmed clusters would lead to a close to fully complete (>99%) cluster catalog.

5.4. Adding information from point-source optimized source identification

So far, we call any noncluster source a contaminant and do not distinguish whether it is a real noncluster source or a spurious (noise) fluctuation. From simulations, we expect that almost all sources in the extent-selected sample have a true underlying astrophysical signal and are not noise fluctuations. The majority of other astrophysical sources fall into the category of point sources, mostly AGN and stars. The source identification for point-like sources is treated separately in a dedicated companion paper (Salvato et al. 2022). The size and variety of the training sample we used to calibrate the point-source identification method generally allows the identification of the counterparts

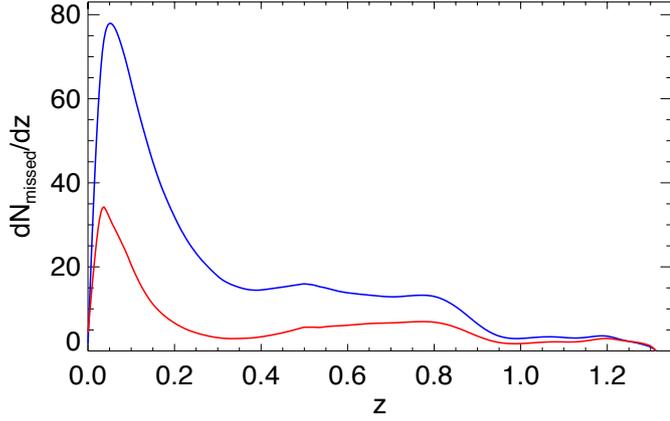


Fig. 18. Redshift distribution of missed systems according to the adopted selection: $f_{\text{cont}} < 0.2$ (blue) and $f_{\text{cont}} < 0.3$ (red)

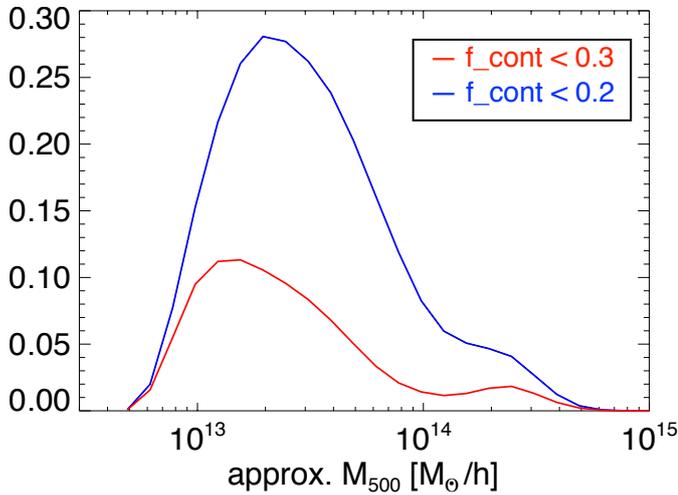


Fig. 19. Estimated M_{500} distribution of missed systems according to the adopted selection: $f_{\text{cont}} < 0.2$ (blue) and $f_{\text{cont}} < 0.3$ (red).

regardless of their nature. Cluster members near the X-ray position can also be identified as the correct counterpart. On the other hand, the lack of a good point like counterpart does not mean that the X-ray emission stems from a cluster. A combination of multiple confused point sources causing the X-ray position to be different from a good counterpart is also possible. Furthermore, combinations of point and extended sources exist as well, for instance, central AGN or miscentered clusters due to a point source on top of a flat surface brightness source. Without high-resolution X-ray imaging, many of these cases cannot be resolved by combining MCMF and point-like identifications.

The primary tool for finding the best point-like counterparts is NWAY (Salvato et al. 2018). NWAY is based on a Bayesian method of finding the best counterparts given a set of photometric priors and the X-ray to optical relative astrometry. In its latest iteration on eFEDS (Salvato et al. 2022), NWAY was complemented with priors in optical to MIR from LS and complemented with *Gaia* proper motion parameters. The tool provides the probabilities that a certain source has an LS counterpart (p_{any}), which is the main output used for purity and completeness calibrations. The thresholds and purity

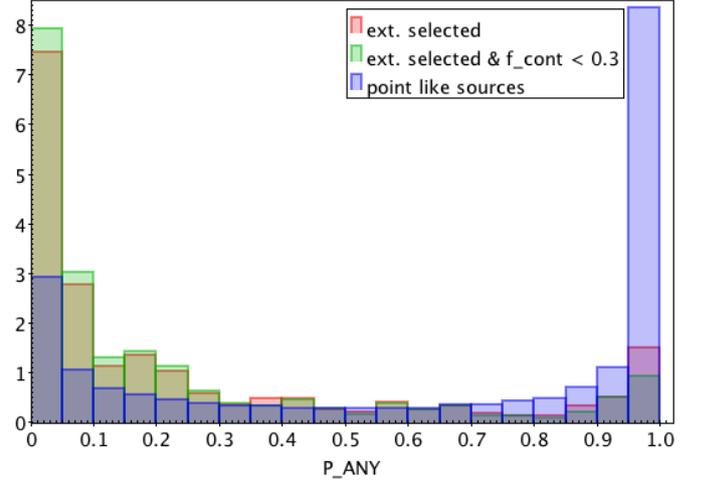


Fig. 20. Normalized distribution of p_{any} for the extent-selected and point-like samples with and without a cut at $f_{\text{cont}} < 0.3$. The distributions differ most for $p_{\text{any}} < 0.3$ and $p_{\text{any}} > 0.7$, reflecting the different mix of X-ray source types in the extent-selected sample with respect to the point-like sample.

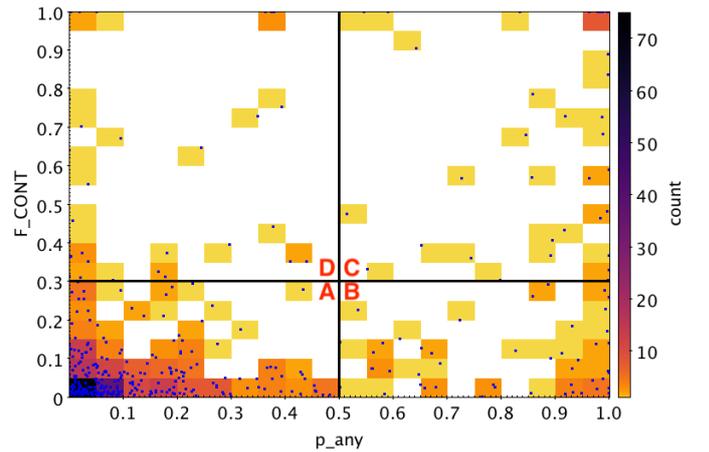


Fig. 21. Distribution of the sample in two confirmation proxies f_{cont} and p_{any} . Black lines split the sources into four sectors (A, B, C, and D). Most sources are at low p_{any} and low f_{cont} (sector A), suggesting that they are good clusters without a good point-like counterpart. The second densest point is at high p_{any} and f_{cont} (sector C), suggesting a good AGN. The bottom right corner (low f_{cont} , high p_{any} , sector B) is the region in which both codes find counterparts and where further reconciliation is required, while in the top left corner (sector D) lie sources without a good counterpart in any of the codes.

derived for the point-source catalog are likely not valid for the extent-selected catalog because p_{any} is directly related to X-ray morphological characteristics.

This difference becomes visible in the different distribution of p_{any} for the point-like and the extent-selected sample shown in Fig. 20.

We therefore chose $p_{\text{any}} = 0.5$ as the boundary between good and less good point-like counterparts. In Fig. 21 we show the distribution of sources in the f_{cont} - p_{any} plane. There is an obvious overdensity at low f_{cont} and low p_{any} , which is the region with the highest probability of being a cluster and with no good point-like counterpart. On the other hand, we find the most likely point sources at high p_{any} and high f_{cont} . Using $f_{\text{cont}} = 0.3$ and $p_{\text{any}} = 0.5$ as division lines, we can split the

plane into four sectors, with rather clear cases for clusters in the lower left and good point sources in the upper right. The top left sector (sector D) is the region in which no method finds a safe counterpart, and in the bottom right sector, both find good counterparts.

Leaving the two sectors with clear counterparts aside we focus first on sector D. This is the sector where we expect to find most of the clusters lost by cleaning, as well as blended AGN or spurious sources. We inspect all 29 sources by eye using the available optical images and the X-ray surface brightness maps. We find a few cases where the source are close to very bright X-ray sources, indicating that those might be spurious sources caused by the bright primary source. We also find four potential cluster matches, although X-ray surface brightness maps do not show a convincing peak at the optical counterpart. From Sect. 5.3 we expect about 8_{-2}^{+2} missing systems in total. In Sect. 5.1 we find four good and three less good counterparts. Even when all these systems are counted as truly missed clusters, the total number of 11 systems would still be within two sigma from our estimate of the incompleteness. Allowing some systems to be projections or AGN dominated will just improve the agreement between observation and prediction.

The second sector of interest is sector B, which contains 74 sources. As already outlined, a high p_{any} by itself does not exclude typical cluster galaxies as possible counterparts. We therefore aim to filter out clear cases in which the best counterpart is a cluster member. To do this, we matched the best counterparts found by NWay with existing spec- z catalogs. Furthermore, we derived red-sequence-based photometric redshifts similar to redMaGiC (Roza et al. 2016) using the LS photometry. For sources for which the red-sequence model is a reasonable fit, we obtain $\sigma_{\Delta z/(1+z)} = 0.014$ for $z_{\text{phot}} < 0.9$ based on 578 spectroscopic redshifts in the joint point source and extended source catalog.

Using maximum offsets of $\Delta z/(1+z) < 0.05$ between MCMF and spec- z and $\Delta z/(1+z) < 0.1$ between MCMF and photo- z , we find 43 systems with consistent redshifts between a point-like counterpart and a cluster candidate. This leaves 31 sources with good counterparts in both catalogs, but with discrepant redshifts. We then visually inspected all systems using HSC, Legacy Survey images, and smoothed X-ray surface brightness maps. In the 43 sources with consistent redshifts, we find one clear and one unclear contaminant (1.3–4.6%). The vast majority are indeed obvious clusters where the X-ray PS counterpart is identical to the BCG.

When we repeated this for the 31 cases with discrepant or invalid redshift estimates, the picture appears different and more complex than for the other subset. Only eight systems appear as clearly unaffected clusters, and nine systems appear point like. Another nine systems appear to be affected by both point-like and extended emission. These systems typically show X-ray emission at the optical cluster position, but the X-ray center seems to be shifted to a location of a good point-like counterpart. Additionally, we find four systems for which a classification is unclear, typically associated with sources of lower p_{any} . To summarize, we find a contamination of 29–74% from counting secure cluster or point sources alone.

Based on these findings, we added a additional flag to the catalog according to the sectors we discussed. The flag corresponding to sector B is split into B1, where PS and cluster redshifts agree, and B2, where the redshift disagrees. This flag can be used to select even cleaner cluster subsamples or to explicitly study AGN that leak into the cluster- and group-candidate catalog.

5.5. Estimate of the contamination in cluster catalog

In addition to the X-ray extent-selected candidate catalog, MCMF was also run on the X-ray point-like source list in eFEDS. More details about the run on point sources are provided in Appendix B. The key difference between the runs is that the richness is estimated from optical information alone and does not include the prior on the aperture that is based on the X-ray count rate within which the richness is extracted. The richness is therefore constructed in a way that it is unbiased against the richness given in the run on extent-selected clusters. With this, the MCMF run on the point-like sample can act as a reference sample for the contaminants because it is dominated by real point-like sources and shares a similar flux distribution as the expected point-source contaminants in the extent-selected sample.

A reference sample for the $f_{\text{cont}} < 0.3$ contaminants can be constructed by applying the same cut of λ versus redshift as was done with the $f_{\text{cont}} < 0.3$ on the extent-selected sample. Second, a flux cut has to be applied to mimic the selection of the extent-selected sample. This was simply done by cutting at the minimum rate we found in the extent-selected sample. To further reduce the number of spurious sources and cluster signal from nearby detected clusters, we excluded 10 arcmin regions around extent-selected clusters and also selected $p_{\text{any}} > 0.5$.

From Fig. 7 we know that the main contamination is expected to be in the low- λ high X-ray based mass corner of the scaling relation. Using the scaling relation derived in Sect. 5.3, we can measure the offset distribution around the scaling relation measured in the sigma of the scatter. In Fig. 22 we show this distribution for the $f_{\text{cont}} < 0.3$ as well as for the full sample. The offset distribution is then a combination of the distribution of true clusters and of point-like contaminants. We therefore aim to fit a model for both populations to derive the fraction of contaminants in the sample. The offset distribution for the reference contaminant sample is very well described by a Gaussian function. We therefore used the best-fit Gaussian function as a model for the contaminant sample, leaving only the normalization as a free parameter. For the clean cluster sample, we used the $f_{\text{cont}} < 0.3$ sample with an additional cut on $p_{\text{any}} < 0.5$ to suppress the majority of contaminating point sources. We then performed an MCMC fit to the observed offset distribution of the $f_{\text{cont}} < 0.3$ sample, with the normalizations of the two subpopulations as the only free parameters. We find a best-fit contamination of $6_{-3}^{+3}\%$, which agrees well with the expected value of 6%. The best-fit model is shown as a red line in Fig. 22.

Motivated by this, we repeat the same exercise for the full sample without an f_{cont} cut. The richness cut for the contaminant sample needs to be changed to a simple $\lambda > 2$ cut, which is the typical lower limit in the default MCMF run. The best-fit model is again shown in Fig. 22. The best-fit contamination is $17_{-3}^{+3}\%$, which fits the expectation of 19.4% from the simulations well and fits our empirical estimate in Sect. 5.2 even better. Compared to Sect. 5.2, our results here do not rely on the simulations, the correct estimate of f_{cont} , or on the completeness.

The main assumption going into the estimate is that the $f_{\text{cont}} < 0.3$ and $p_{\text{any}} < 0.5$ (class A) subsample is close to a clean sample and that the p_{any} cut does not introduce an additional selection affecting the offset distribution in addition to increasing purity. To test this, we repeated the population modeling on the $f_{\text{cont}} < 0.3$ and $p_{\text{any}} > 0.5$ sample. If the assumption is true that the selected subsample is indeed clean, then the expected 6% contaminants should be inside this sample. The result is shown in Fig. 22. Although this subsample is small, the composite model does not appear to match the data as well as

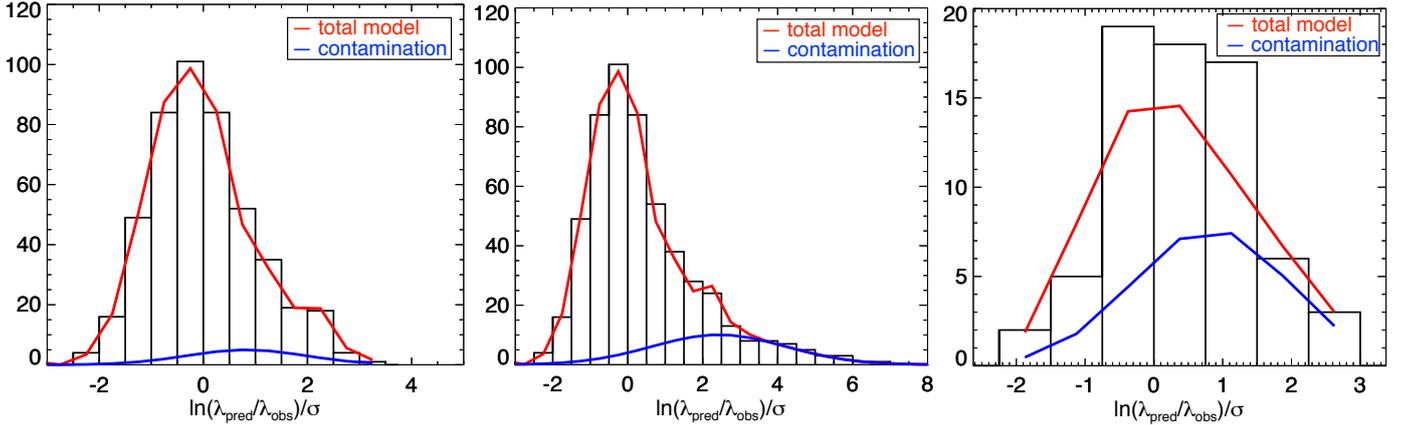


Fig. 22. Offset distribution around the best-fit scaling relation in units of the standard deviation σ for the $f_{\text{cont}} < 0.3$ (left), the full (middle), and the $f_{\text{cont}} < 0.3$ & $p_{\text{any}} > 0.5$ sample (right). The red line shows the best-fit total model, and the blue line shows the model of the contaminating population.

in the previous cases. This is likely an indication that the typical scaling of the cluster population is different from the assumed model. One reason for this might be that this sample could host more relaxed cool-core clusters than the clean reference sample. Cool-core clusters usually show small X-ray to BCG offsets so that NWAY might identify the BCG as a good point-source candidate. Another potential reason might be the significant contribution of a point source to the cluster emission. Both effects would shift the peak of the distribution to higher values, as suggested in the observed distribution. Leaving these issues aside, the fit suggests a significantly higher level of contamination than the other samples. The total number of contaminants is estimated to be 28.4 ± 3.9 , which again is similar to the expected 28 systems from assuming 6% contamination of the overall $f_{\text{cont}} < 0.3$ sample. Although this number needs to be interpreted with care, it at least suggests that the majority of contaminants in the sample are indeed at $p_{\text{any}} > 0.5$.

6. Conclusions

The eROSITA eFEDS field is observed to a depth similar to what is expected to be reached in this region at the end of the full 4-yr all-sky survey. In this paper we described the optical identification of X-ray selected galaxy cluster and group candidates from this field. The optical identification yields a cluster catalog that reaches $z = 1.3$ and confirms groups out to $z \sim 0.3$. We confirmed and obtained redshifts for $>98\%$ of the true clusters and groups using optical photometric data alone while simultaneously reducing the candidate catalog contamination by 70%. Using the richness to X-ray mass proxy scaling relation and the X-ray source candidates, we predicted the incompleteness induced by optical identification to be 2% (5%) for the $f_{\text{cont}} < 0.3$ (0.2) optical confirmation thresholds. The modeled incompleteness is consistent with the number of systems found by matching the eFEDS catalog to various cluster and group catalogs available in the literature. The recovered fraction of confirmed systems as well as its dependence on source extent likelihood agrees well with the predictions from dedicated simulations of the eFEDS field. The catalog contamination is estimated by modeling the impact of point-like sources on the scatter distribution around the derived scaling relation. The final confirmed cluster catalog contamination of $6 \pm 3\%$ agrees well with the expectation from the adopted cut in f_{cont} and the

expected initial level of contamination. By adding the systems that were found to be missed by MCMF, we expect the catalog to include $>99\%$ of the real clusters and groups in the X-ray cluster candidate list. In addition to optical information related to cluster confirmation, we also provide optical estimators of the dynamical state of clusters and show its application to selecting galaxy cluster mergers.

These results provide a positive outlook for the future eROSITA all-sky surveys because we were able to show that the vast majority of clusters and groups can be confirmed with good-quality photometric data alone. By complementing the optical follow-up with a cross-comparison to existing low-redshift group catalogs, it should be possible to reach high completeness in the optical confirmation of cluster candidates.

Acknowledgements. This work is based on data from eROSITA, the soft X-ray instrument aboard SRG, a joint Russian-German science mission supported by the Russian Space Agency (Roskosmos), in the interests of the Russian Academy of Sciences represented by its Space Research Institute (IKI), and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The SRG spacecraft was built by Lavochkin Association (NPOL) and its subcontractors, and is operated by NPOL with support from the Max Planck Institute for Extraterrestrial Physics (MPE). The development and construction of the eROSITA X-ray instrument was led by MPE, with contributions from the Dr. Karl Remeis Observatory Bamberg and ECAP (FAU Erlangen-Nuernberg), the University of Hamburg Observatory, the Leibniz Institute for Astrophysics Potsdam (AIP), and the Institute for Astronomy and Astrophysics of the University of Tuebingen, with the support of DLR and the Max Planck Society. The Argelander Institute for Astronomy of the University of Bonn and the Ludwig Maximilians Universitaet Munich also participated in the science preparation for eROSITA. The eROSITA data shown here were processed using the eSASS/NRTA software system developed by the German eROSITA consortium. The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University. This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at <http://dm.lsst.org>. This work was supported in part by the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and JSPS KAKENHI Grant Nos. JP18K03693 and JP19KK0076. In addition, we acknowledge support from the Excellence Cluster ORIGINS, the Max Planck Society Faculty Fellowship program and the Ludwig-Maximilians-Universitaet. This paper is based on

data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z -band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Cientifico e Tecnologico and the Ministerio da Ciencia, Tecnologia e Inovacao, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenossische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l'Espai (IEEC/CSIC), the Institut de Fisica d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universitat Munchen and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program "The Emergence of Cosmological Structures" Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant # 114A11KY5B20160057), and Chinese National Natural Science Foundation (Grant # 11433005). The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration. The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO. JW acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 -390783311.

References

- Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2020, *Phys. Rev. D*, **102**, 023509
- Adami, C., Giles, P., Koulouridis, E., et al. 2018, *A&A*, **620**, A5
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, *PASJ*, **70**, S4
- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018b, *PASJ*, **70**, S8
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, *PASJ*, **71**, 114
- Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, **49**, 409
- Bahar, Y. E., Bulbul, E., Clerc, N., et al. 2022, *A&A*, **661**, A7 (eROSITA EDR SI)
- Banerjee, P., Szabo, T., Pierpaoli, E., et al. 2018, *New Astron.*, **58**, 61
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Blanton, M. R., Bershad, M. A., Abolfathi, B., et al. 2017, *AJ*, **154**, 28
- Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, *ApJ*, **878**, 55
- Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, **70**, S5
- Brunner, H., Liu, T., Lamer, G., et al. 2022, *A&A*, **661**, A1 (eROSITA EDR SI)
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Bulbul, E., Chiu, I.-N., Mohr, J. J., et al. 2019, *ApJ*, **871**, 50
- Chiu, I. N., Okumura, T., Oguri, M., et al. 2020a, *MNRAS*, **498**, 2030
- Chiu, I. N., Umetsu, K., Murata, R., Medezinski, E., & Oguri, M. 2020b, *MNRAS*, **495**, 428
- Chiu, I.-N., Ghirardini, V., Liu, A., et al. 2022, *A&A*, **661**, A11 (eROSITA EDR SI)
- Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, *ApJ*, **648**, L109
- Comparat, J., Eckert, D., Finoguenov, A., et al. 2020, *Open J. Astrophys.*, **3**, 13
- Coupon, J., Czakon, N., Bosch, J., et al. 2018, *PASJ*, **70**, S7
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, **157**, 168
- Dressler, A. 1980, *ApJ*, **236**, 351
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astron. Geophys.*, **50**, 5.12
- Finoguenov, A., Guzzo, L., Hasinger, G., et al. 2007, *ApJS*, **172**, 182
- Finoguenov, A., Rykoff, E., Clerc, N., et al. 2020, *A&A*, **638**, A114
- Ghirardini, V., Bahar, Y. E., Bulbul, E., et al. 2022, *A&A*, **661**, A12 (eROSITA EDR SI)
- Gladders, M. D., & Yee, H. K. C. 2000, *AJ*, **120**, 2148
- Grandis, S., Klein, M., Mohr, J. J., et al. 2020, *MNRAS*, **498**, 771
- Harvey, D., Massey, R., Kitching, T., Taylor, A., & Tittley, E. 2015, *Science*, **347**, 1462
- Hilton, M., Sifón, C., Naess, S., et al. 2021, *ApJS*, **253**, 3
- Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *ApJS*, **199**, 26
- Ider Chitham, J., Comparat, J., Finoguenov, A., et al. 2020, *MNRAS*, **499**, 4768
- King, I. 1962, *AJ*, **67**, 471
- Klein, M., Mohr, J. J., Desai, S., et al. 2018, *MNRAS*, **474**, 3324
- Klein, M., Grandis, S., Mohr, J. J., et al. 2019, *MNRAS*, **488**, 739
- Lang, D. 2014, *AJ*, **147**, 108
- Lang, D., Hogg, D. W., & Mykytyn, D. 2016, *Astrophysics Source Code Library* [record ascl:1604.008]
- Leauthaud, A., Finoguenov, A., Kneib, J.-P., et al. 2010, *ApJ*, **709**, 97
- Liu, A., Bulbul, E., Ghirardini, V., et al. 2022a, *A&A*, **661**, A2 (eROSITA EDR SI)
- Liu, T., Merloni, A., Comparat, J., et al. 2022b, *A&A*, **661**, A27 (eROSITA EDR SI)
- Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010, *MNRAS*, **406**, 1759
- Meisner, A. M., Lang, D., Schlafly, E. F., & Schlegel, D. J. 2019, *PASP*, **131**, 124504
- Merloni, A., Predehl, P., Becker, W., et al. 2012, *ArXiv e-prints* [arXiv:1209.3114]
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, *PASJ*, **70**, S1
- Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, *Nature*, **379**, 613
- Murata, R., Oguri, M., Nishimichi, T., et al. 2019, *PASJ*, **71**, 107
- Oguri, M. 2014, *MNRAS*, **444**, 147
- Oguri, M., Lin, Y.-T., Lin, S.-C., et al. 2018, *PASJ*, **70**, S20
- Okabe, N., Oguri, M., Akamatsu, H., et al. 2019, *PASJ*, **71**, 79
- Pacaud, F., Clerc, N., Giles, P. A., et al. 2016, *A&A*, **592**, A2
- Piffaretti, R., Arnaud, M., Pratt, G. W., Pointecouteau, E., & Melin, J. B. 2011, *A&A*, **534**, A109
- Planck Collaboration XXIV. 2016, *A&A*, **594**, A24
- Planck Collaboration XXVII. 2016, *A&A*, **594**, A27
- Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, *A&A*, **647**, A1
- Rosati, P., Borgani, S., & Norman, C. 2002, *ARA&A*, **40**, 539
- Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, *ApJ*, **708**, 645
- Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, *MNRAS*, **461**, 1431
- Rykoff, E. S., Rozo, E., Hollowood, D., et al. 2016, *ApJS*, **224**, 1
- Salvato, M., Buchner, J., Budavári, T., et al. 2018, *MNRAS*, **473**, 4937
- Salvato, M., Wolf, J., Dwelly, T., et al. 2022, *A&A*, **661**, A3 (eROSITA EDR SI)
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525

- Tempel, E., Tamm, A., Gramann, M., et al. 2014, *A&A*, **566**, A1
- Truemper, J. 1982, *Adv. Space Res.*, **2**, 241
- Tully, R. B. 2015, *AJ*, **149**, 171
- van Weeren, R. J., Röttgering, H. J. A., Brügger, M., & Hoefl, M. 2010, *Science*, **330**, 347
- Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, *ApJ*, **692**, 1060
- Voges, W., Aschenbach, B., Boller, T., et al. 1999, *A&A*, **349**, 389
- Wen, Z. L., & Han, J. L. 2013, *MNRAS*, **436**, 275
- Wen, Z. L., & Han, J. L. 2015, *ApJ*, **807**, 178
- Willis, J. P., Oguri, M., Ramos-Ceja, M. E., et al. 2021, *MNRAS*, **503**, 5624
- ⁸ Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), 11F of AS/NTU Astronomy-Mathematics Building, No.1, Sec. 4, Roosevelt Rd, Taipei 10617, Taiwan
- ⁹ IRAP, Université de Toulouse, CNRS, UPS, CNES, Toulouse, France
- ¹⁰ Kobayashi-Maskawa Institute for the Origin of Particles and the Universe (KMI), Nagoya University, Nagoya 464-8602, Japan
- ¹¹ National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
- ¹² SOKENDAI (The Graduate University for Advanced Studies), Mitaka, Tokyo 181-8588, Japan
- ¹³ Physics Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
- ¹⁴ Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
- ¹⁵ Core Research for Energetic Universe, Hiroshima University, 1-3-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
- ¹⁶ Argelander-Institut für Astronomie (AIfA), Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
- ¹⁷ Department of Physics, Nara Women's University, Kitaoyanishimachi, Nara 630-8506, Japan
- ¹⁸ CRAR, The University of Western Australia, 7 Fairway, Crawley WA 6009, Australia
- ¹⁹ Exzellenzcluster ORIGINS, Boltzmannstr. 2, 85748 Garching, Germany
-
- ¹ Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany
e-mail: matthias.klein@physik.lmu.de
- ² Research Center for the Early Universe, University of Tokyo, Tokyo 113-0033, Japan
- ³ Department of Physics, University of Tokyo, Tokyo 113-0033, Japan
- ⁴ Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, Chiba 277-8582, Japan
e-mail: masamune.oguri@ipmu.jp
- ⁵ Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse 1, 85748 Garching, Germany
- ⁶ Tsung-Dao Lee Institute, and Key Laboratory for Particle Physics, Astrophysics and Cosmology, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, PR China
- ⁷ Department of Astronomy, School of Physics and Astronomy, and Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai 200240, PR China

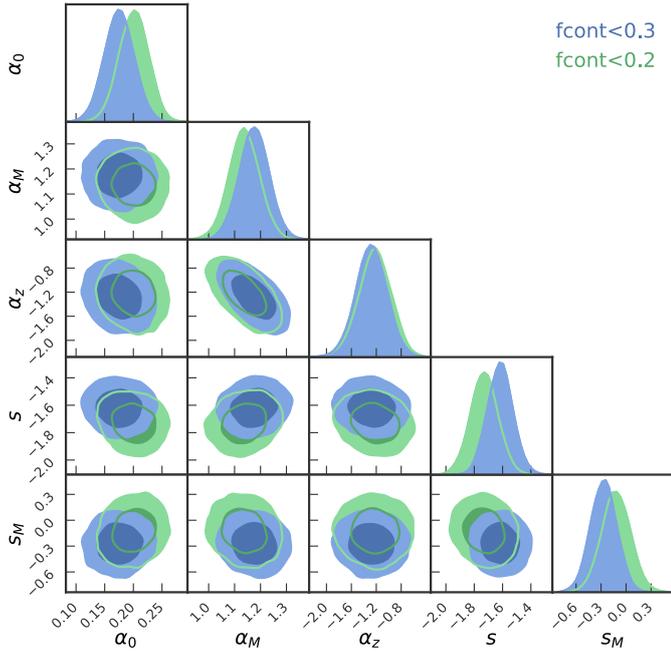


Fig. A.1. Plots of the posteriors for the parameters of the richness–mass scaling relation fit using the sample with $f_{\text{cont}} < 0.2$ (0.3) in green (blue). The two sets of posteriors are consistent. We expect that the minor differences are due to the larger number of contaminating random line-of-sight superpositions in the $f_{\text{cont}} < 0.3$ case.

Appendix A: Richness–mass scaling relation

In this section we provide some further details and results on the scaling relation used in Sect. 5.3. To fit the free parameters of the scaling relation, we set up a likelihood for each cluster i to conform with the assumed relation,

$$\mathcal{L}_i = C_i^{-1} P(\ln \lambda_i | M_{500,i}, z_i), \quad (\text{A.1})$$

where $(\lambda_i, M_{500,i}, z_i)$ are the richness, mass, and redshift of the cluster, respectively. The normalization C_i is required to ensure that the likelihood is normalized over the range of possible richnesses. We recall that the likelihood is the probability of data (in this case λ_i) given the model (here $(M_{500,i}, z_i, \alpha_0, \alpha_M, \alpha_z, s, \text{ and } S_M)$). It therefore needs to be normalized over all possible data. In our case, this means that the richness has to be larger than the richness cut associated with the optical cleaning, that is, $\lambda > \lambda_{\text{cut}}(z_i)$. Thus,

$$C_i = \int_{\ln \lambda_{\text{cut}}(z_i)}^{\text{inf}} d \ln \lambda P(\ln \lambda | M_{500,i}, z_i). \quad (\text{A.2})$$

This term ensures that the incompleteness induced by the optical cleaning is properly accounted for.

Moreover, the X-ray selection function does not impact this calculation. M_{500} is an analytical function of the measured count rate. Any X-ray selection can thus be expressed as some function $\mathcal{S}_X(M_{500}, z)$. If we explicitly account for the X-ray selection function, it appears in both the numerator and the denominator (normalization condition) of Eq. A.1, and therefore the selection cancels out.

Assuming a wide, flat prior for all parameters, we sampled a posterior of the scaling relation parameters for the sample selected by $f_{\text{cont}} < 0.2$ (0.3), with the additional redshift cut of $z < 1.2$. The marginal contours of the posterior plot are shown in

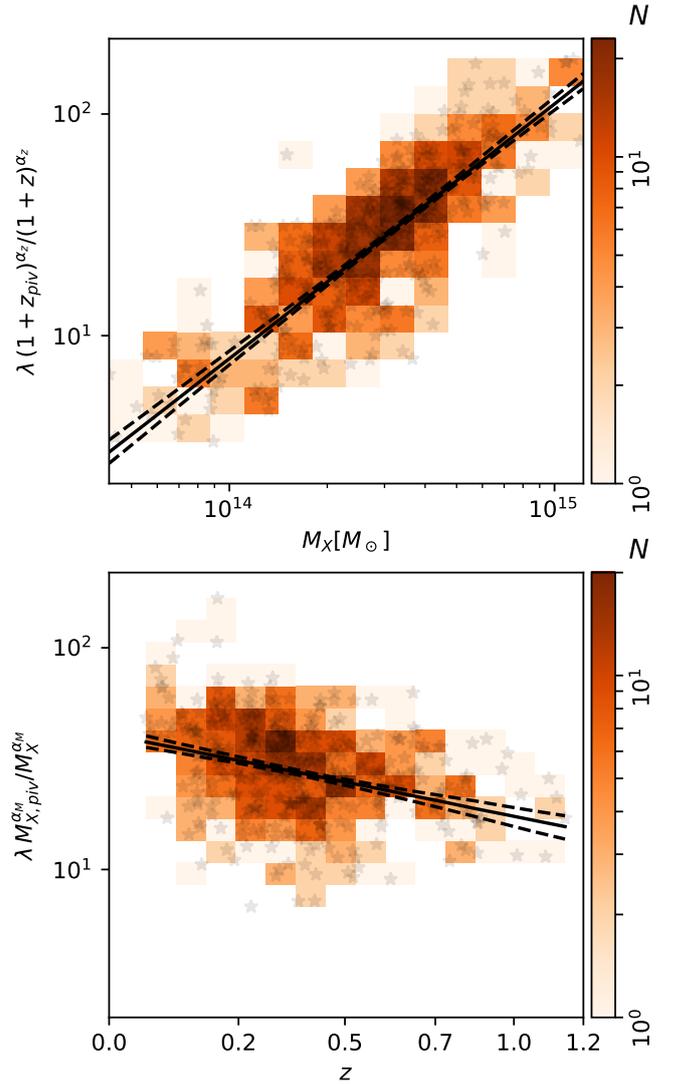


Fig. A.2. *Upper panel:* Redshift-trend-corrected richness vs. mass for the $f_{\text{cont}} < 0.2$. The best-fit mass trend (solid black) and 1σ uncertainty (dashed black) are shown. *Lower panel:* Mass-trend-corrected richness vs. redshift for the $f_{\text{cont}} < 0.2$ sample with best-fit redshift trend and uncertainty, as above.

Fig. A.1. The $f_{\text{cont}} < 0.2$ sample is by construction a subset of the $f_{\text{cont}} < 0.3$ sample. The two posteriors are therefore far from being independent. The difference between the posteriors is much smaller than 1σ . We interpret the small difference to be due to the residual contamination that is included by the more lenient $f_{\text{cont}} < 0.3$ cut. For this reason, we use the posterior derived from the $f_{\text{cont}} < 0.2$ sample below.

To investigate the adequacy of our best-fit mass and redshift trends, we plot in Fig. A.2 the mass- and redshift-trend-corrected richnesses to highlight the redshift and mass trend. We also overplot the best-fit mass and redshift trends. They match the data well, showing the adequacy of our best fit. The mass proxy we used is a modified count rate. From our eFEDS simulations, we know that measured count rates are biased with respect to true count rates, where the bias depends on the input count rate, size, and signal-to-noise ratio. This might bias the redshift trend. Preliminary results using weak gravitational lensing (Chiu et al. 2022) suggest no redshift evolution of the lambda-mass relation.

Appendix B: MCMF X-ray point-like sources

The MCMF run on the point like X-ray candidate catalog was performed in two steps. First, MCMF was run in normal mode using the X-ray source counts as a mass proxy. To avoid unrealistically large apertures, the maximum count rate was limited to that of a typical $\text{ext_like}=50$ source. True clusters brighter than this would very likely be detected as extended sources in eFEDS. The main aim of this MCMF run was to provide redshifts for possible counterparts. In the second step, these redshifts and the X-ray positions were used to calculate richnesses. Similar to the approach in redMaPPer (Rykoff et al. 2016), we defined our richness through a scaling relation between aperture r_λ and the optical only based richness λ_{OPT} as

$$\lg r_\lambda = (\lg \lambda_{\text{OPT}})m + b, \quad (\text{B.1})$$

with a slope m and a normalization b . Our goal here was to calibrate our optical only richness to the richness obtained using X-ray priors that use a proxy of r_{500} . We therefore simply fit for this relation using the MCMF results on the extent-selected sample and find $m = 0.29$ and $b = -0.496$ as suitable parameters to define our relation. Due to the high point-source density and source-splitting of extended sources, there is a significant overlap between the extent-selected sample and the point-source selected sample. We find 206 matches within 2 arcminutes and within a redshift offset of $\Delta z < 0.04$. The richnesses of the optical-only versus the default MCMF richnesses are shown in the left panel of Fig. B.1. We also matched the point-source sample with the SDSS redMaPPer (Rykoff et al. 2016) sample using a 90-arcsecond search radius and the same maximum redshift offset as before. The comparison to this sample is shown in the right panel of Fig. B.1. Incompleteness in SDSS starts to impact the richness estimate in redMaPPer for $z > 0.4$, causing increased scatter and a shift in the richness compared to the MCMF optical-only richness λ_{OPT} . For redshifts below $z = 0.4$, the richnesses of the methods appear to agree.

Appendix C: Additional results for the dynamical state estimators

In Fig. C.1 we show the correlation between different estimators that probe the cluster morphology. All estimators are based on the galaxy density maps and are therefore prone to similar systematics such as projections, masks, or other artifacts. The estimators themselves probe different properties that are associated with unrelaxed systems. The three Wen & Han (2013) based estimators (Greek letters) focus on deviation from symmetry or from the adopted 2D model (see Sect. 3.1.5). The ellipticity and the center shift with respect to the X-ray position are mostly independent of the first set of estimators because they also trigger on perfectly symmetric well-modeled clusters, but with unusual shape or offset. Estimates become more noisy and therefore less reliable with lower richness and higher redshifts. We therefore suggest that these estimators are most useful when a rather high richness threshold is adopted, for example, $\lambda > 50$, and is restricted to redshifts $z < 0.8$.

In Table C.1 we show the 25 most disturbed eFEDS clusters with $\lambda > 50$ according to the sum of the estimators α , δ , β , ellipticity, and center shift, called S_{DYN} . We further provide a flag if another eFEDS cluster is nearby. The flag is 1 if another eFEDS cluster is nearby, 2 if the nearby cluster has a higher disturbance estimate, and 3 if there are two other nearby clusters with higher disturbance estimates. With this, Table C.1 shows 21 individual merging systems, where two systems appear as three

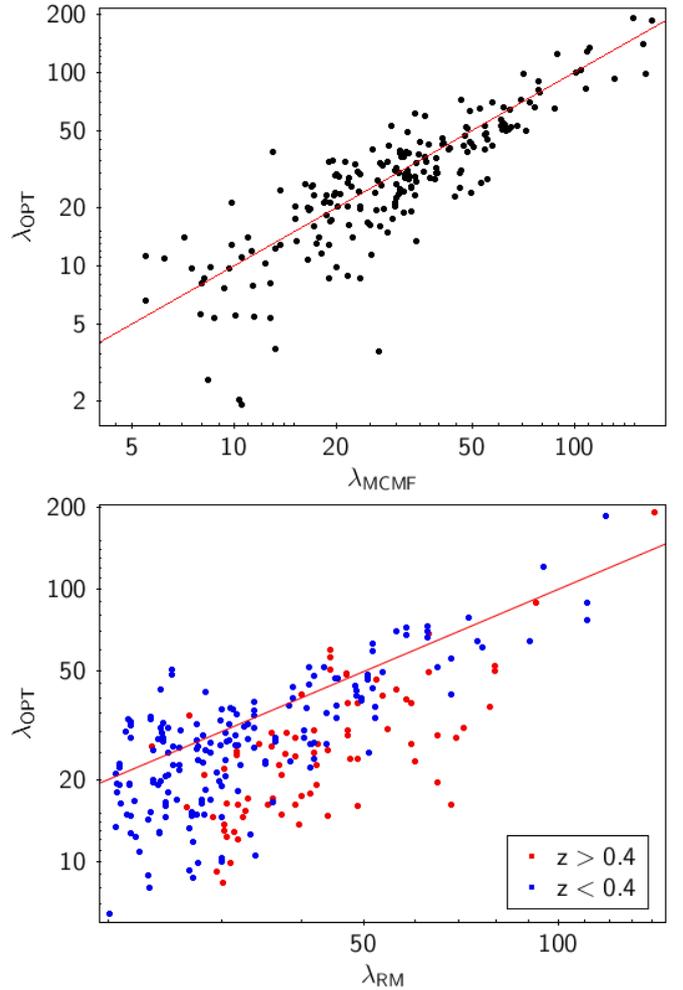


Fig. B.1. Top: Comparison between the optical only estimated richness λ_{OPT} and the default MCMF richness (which uses X-ray prior). Bottom: Similar, but comparing to richnesses from matches to the SDSS redMaPPer sample. The red line indicates the one-to-one relation.

eFEDS sources in the table. Furthermore, the table shows that 10 systems have at least one other eFEDS cluster nearby.

Table C.2 shows cluster pairs and potential mergers using the requirement of having another optical and X-ray detection within $2.5 * r_{500}$, as suggested in Sect. 4.3. The list is sorted by S_{DYN} . We find that 29 eFEDS clusters fulfill the selection. The list contains 17 individual systems.

Appendix D: Column description of the results tables

Because the tables containing the main results are too long and include too many columns, we provide them only in electronic form on the CDS and the official eROSITA website. To give at least a brief overview of the available information, we provide here a short description of the key entries of these tables.

In Table D.1 we list the key entries of the results table. This includes the most essential entries from the X-ray catalog that were used in this work. Furthermore, we list entries related to the best optical counterpart based on the combination of the MCMF runs on HSC and LS. We note here that similar entries also exist for the second-best optical counterpart, which contains "2BEST" in their column names.

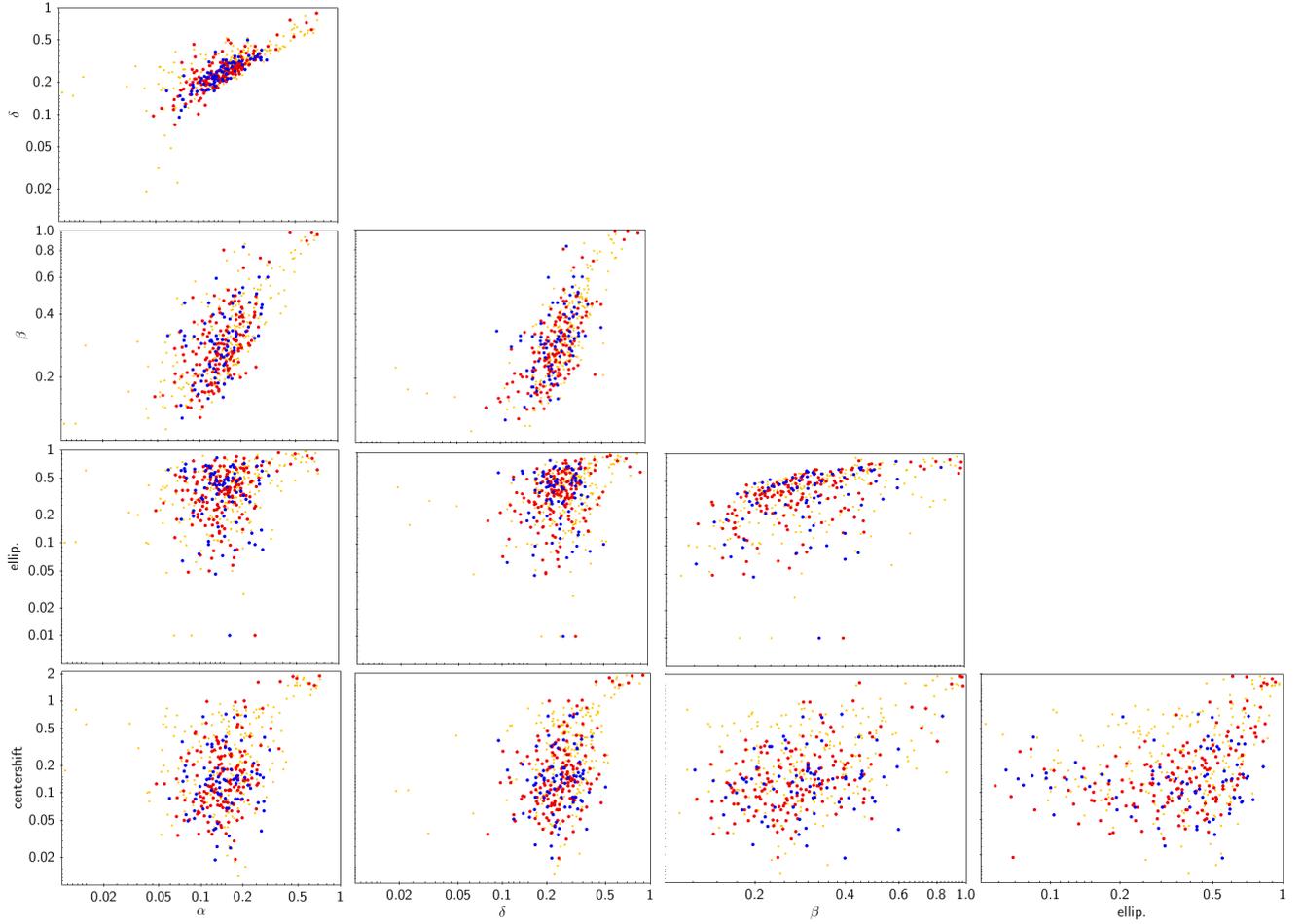


Fig. C.1. Comparison of the five different optical estimators of cluster dynamical state that are based on model fitting to the red-sequence galaxy density map. The same richness thresholds as in Fig. C.1 are color-coded: yellow for $\lambda < 25$, red for $25 < \lambda < 50$, and blue for $\lambda > 50$.

In Table D.2 we list columns provided in an auxiliary table containing the results for the MCMF runs on the HSC and LS datasets.

Table C.1. Top 25 most unrelaxed eFEDS clusters with $\lambda > 50$, according to S_{DYN} .

NAME	RA	DEC	z	λ	α	δ	β	Ellip.	cent. shift	S_{DYN}	Flag
eFEDS J093431.3-002309	143.630	-0.386	0.335	54.6	0.28	0.34	1.04	0.75	0.26	2.67	1
eFEDS J092220.4+034806	140.585	3.802	0.269	51.9	0.21	0.29	0.84	0.71	0.41	2.46	1
eFEDS J085620.7+014649	134.086	1.780	0.732	70.9	0.11	0.23	1.04	0.69	0.32	2.38	1
eFEDS J091302.1+035000	138.259	3.834	0.455	59.3	0.31	0.32	0.60	0.53	0.31	2.07	1
eFEDS J084910.6+024117	132.294	2.688	0.830	68.0	0.13	0.21	0.59	0.82	0.29	2.05	
eFEDS J092209.3+034628	140.539	3.775	0.268	95.6	0.26	0.35	0.50	0.67	0.26	2.04	2
eFEDS J084823.2+041205	132.097	4.201	0.872	90.8	0.24	0.33	1.04	0.13	0.29	2.03	
eFEDS J092046.2+002849	140.193	0.480	0.400	55.5	0.27	0.37	0.60	0.43	0.26	1.92	
eFEDS J084459.2-011903	131.247	-1.317	0.447	53.1	0.28	0.32	0.43	0.59	0.26	1.89	2
eFEDS J091305.9+035021	138.275	3.839	0.454	93.9	0.18	0.30	0.49	0.65	0.27	1.89	
eFEDS J090328.7-013622	135.870	-1.606	0.443	72.3	0.17	0.28	0.45	0.70	0.28	1.87	
eFEDS J092202.2+034520	140.510	3.756	0.268	59.6	0.20	0.34	0.51	0.50	0.33	1.87	3
eFEDS J092339.0+052654	140.913	5.449	0.373	56.1	0.17	0.30	0.38	0.67	0.26	1.78	
eFEDS J083933.8-014044	129.891	-1.679	0.279	78.6	0.12	0.25	0.45	0.67	0.28	1.77	1
eFEDS J084637.1-002256	131.655	-0.382	0.293	52.7	0.08	0.23	0.45	0.71	0.28	1.75	
eFEDS J091315.0+034850	138.313	3.814	0.444	88.8	0.15	0.32	0.39	0.43	0.42	1.71	3
eFEDS J084223.1+003340	130.596	0.561	1.077	52.8	0.17	0.30	0.33	0.50	0.41	1.71	
eFEDS J083125.9+015533	127.858	1.926	0.684	61.6	0.12	0.31	0.30	0.65	0.32	1.69	1
eFEDS J082820.5-000721	127.086	-0.123	0.845	61.5	0.27	0.35	0.32	0.47	0.28	1.68	1
eFEDS J092049.5+024513	140.206	2.754	0.285	79.0	0.21	0.32	0.54	0.32	0.28	1.66	1
eFEDS J090915.3-010104	137.314	-1.018	0.822	54.6	0.23	0.35	0.45	0.28	0.28	1.60	
eFEDS J092212.0-002731	140.550	-0.459	0.318	108.1	0.18	0.31	0.27	0.55	0.27	1.59	1
eFEDS J084246.9-000917	130.696	-0.155	0.415	61.3	0.29	0.40	0.44	0.08	0.33	1.55	
eFEDS J093207.5-021317	143.032	-2.221	0.666	57.5	0.11	0.19	0.36	0.61	0.28	1.55	
eFEDS J090129.2-013853	135.372	-1.648	0.318	104.7	0.20	0.31	0.41	0.36	0.27	1.55	1

Notes. The table is sorted by the value S_{DYN} , the linear combination of the dynamical state estimators α , δ , β , ellipticity, and center shift. It lists the 25 eFEDS clusters with $f_{\text{cont}} < 0.3$ and $\lambda > 50$. Meaning of the flag values: 1) at least one other eFEDS source is nearby, 2) another eFEDS source with $\lambda > 50$ is nearby and has higher S_{DYN} . 3) same as 2), but having two eFEDS clusters with higher S_{DYN} . The positions are given in the J2000 system.

Table C.2. Close cluster pairs or mergers, selected by having an optical and X-ray neighbor within $2.5r_{500}$.

NAME	RA	DEC	z	λ	D_{OPT} (R_{500})	D_{X} (R_{500})	S_{DYN}	Flag
eFEDS J085541.2+002740	133.922	0.461	0.156	8.7	0.94	1.63	3.59	1
eFEDS J091415.0+022709	138.562	2.453	0.330	19.8	0.66	0.89	3.09	1
eFEDS J090146.2-013756	135.443	-1.632	0.304	38.5	1.39	1.23	2.77	1
eFEDS J091354.7+025323	138.478	2.890	0.423	24.5	0.87	1.55	2.69	1
eFEDS J092220.4+034806	140.585	3.802	0.269	51.9	2.12	0.90	2.46	1
eFEDS J085620.7+014649	134.086	1.780	0.732	70.9	1.93	1.89	2.38	1
eFEDS J083921.0-014149	129.838	-1.697	0.278	39.5	0.93	0.84	2.19	1
eFEDS J091302.1+035000	138.259	3.834	0.455	59.3	1.06	0.37	2.07	1
eFEDS J092209.3+034628	140.539	3.775	0.268	95.6	2.05	0.43	2.04	2
eFEDS J090137.7+030253	135.407	3.048	0.188	16.4	1.23	0.68	2.01	1
eFEDS J091305.9+035021	138.275	3.839	0.454	93.9	0.99	0.34	1.89	2
eFEDS J092202.2+034520	140.510	3.756	0.268	59.6	0.91	0.55	1.87	3
eFEDS J091213.4-021621	138.056	-2.273	0.160	25.5	1.28	0.29	1.80	1
eFEDS J083933.8-014044	129.891	-1.679	0.279	78.6	0.86	0.64	1.77	2
eFEDS J091851.7+021432	139.716	2.242	0.280	28.4	1.15	0.82	1.76	1
eFEDS J091315.0+034850	138.313	3.814	0.444	88.8	0.78	0.83	1.71	3
eFEDS J083930.3-014348	129.876	-1.730	0.271	11.3	0.59	1.03	1.70	3
eFEDS J091358.1+025707	138.492	2.952	0.435	19.0	1.95	1.38	1.63	2
eFEDS J090131.1+030056	135.380	3.016	0.194	62.3	0.47	0.43	1.53	2
eFEDS J090750.1+025006	136.959	2.835	0.648	16.0	1.56	1.51	1.53	1
eFEDS J085751.6+031039	134.465	3.178	0.198	97.9	1.70	2.13	1.52	1
eFEDS J083806.9-003600	129.529	-0.600	0.434	18.0	1.15	0.71	1.43	1
eFEDS J085436.6+003835	133.653	0.643	0.110	31.8	0.74	0.24	1.41	1
eFEDS J093009.0+040144	142.538	4.029	0.342	21.6	1.44	1.54	1.39	1
eFEDS J093003.3+035630	142.514	3.942	0.330	34.9	1.43	1.45	1.38	2
eFEDS J093513.0+004757	143.805	0.799	0.364	169.0	1.45	1.55	1.35	1
eFEDS J085433.0+004009	133.638	0.669	0.113	20.4	1.05	0.35	1.28	2
eFEDS J085627.2+014217	134.113	1.705	0.732	100.5	1.08	1.67	1.26	2
eFEDS J093500.7+005417	143.753	0.905	0.381	54.8	1.93	2.04	1.26	2

Notes. The table is sorted by the value S_{DYN} , the linear combination of the dynamical state estimators α , δ , β , ellipticity, and center shift. Beside $f_{\text{cont}} < 0.3$, the list is selected by the distance to the next optical and X-ray to be $D_{\text{OPT}} < 2.5$ and $D_{\text{X}} < 2.5$. Flags are the same as in Table C.1. The positions are given in the J2000 system.

Table D.1. Column names and description of the entries of the main results table.

Column name	Description
Name	eROSITA source name
ID_SRC	Source ID of detection pipeline
RA_CORR	RA of the X-ray center
DEC_CORR	DEC of the X-ray center
EXT	Source extent in X-ray
EXT_ERR	Error in source extent
EXT_LIKE	Extent likelihood
DET_LIKE_0	Detection likelihood
ML_RATE_0	X-ray count rate
ML_RATE_ERR_0	error on count rate
F_CONT_BEST_COMB	f_{cont} of best optical counter part: -1, -2 for manually added systems
Z_BEST_COMB	Photometric redshift of best counter part
SIGMA_Z_BEST_COMB	Uncertainty of the photometric redshift
LAMBDA_BEST_COMB	Richness of best counter part
ELAMBDA_BEST_COMB	Uncertainty richness
SURV_BEST_COMB	Survey used: -1 group catalog, 1 LS, 2 HSC
RA_OPTCEN_BEST_COMB	Ra of optical center of best counter part
DEC_OPTCEN_BEST_COMB	Same but DEC
SPEC_Z_BEST_COMB	Spec-z
N_SPEC	Number of redshifts used for spec-z
MASSPROX_BEST_COMB	Approx. M500 used for richness estimate [1/h]
FLAG_OPTICAL_X_POS	Value of the optical footprint map at X-ray position
MASKFRAC_3_FOOT	Fraction of the area not int the optical footprint within a 3 arcmin radiuss
MASKFRAC_3_FLAGGED	Similar but including flagged area due to bright star masks
MASKFRAC_R500_FOOT_BEST_COMB	Same as MASKFRAC_3_FOOT but for a r_500 sized region
MASKFRAC_R500_FLAGGED_BEST_COMB	Same as MASKFRAC_3_FLAGGED but for a r_500 sized region
ALPHA	Dynamical state estimator alpha
BETA	Dynamical state estimator beta
DELTA	Dynamical state estimator delta
ELLIP	Dynamical state estimator: model ellipticity
CENTERSHIFT	Shift of the centre of the 2D density model with respect to X-ray position
DIST_NEXT_OPT	Distance to next optical structure in the galaxy density map
DIST_NEXT_XCLUST	Distance to next extent-selected eFEDS cluster
MCMF_NWAY_SECTOR	Sector in f_{cont} - p_{any} plane according to Sect. 5.4

Table D.2. Column names and descriptions for MCMF individual results on LS and HSC

Column name	Description
F_CONT_LFC	f_{cont} of best optical counter part of individual MCMF run
Z_LFC	Corresponding redshift
LAMBDA_LFC	Corresponding richness
ELAMBDA_LFC	Uncertainty of the richness
Z_P1	Redshift of first peak in λ versus redshift plot
Z_P2	Same for second peak
Z_P3	Same for third peak
LAMBDA_P1	Richness of first peak
LAMBDA_P2	Richness of second peak
LAMBDA_P3	Richness of third peak
ELAMBDA_P1	Uncertainty on the richness of first peak
ELAMBDA_P2	Uncertainty on the richness of second peak
ELAMBDA_P3	Uncertainty on the richness of third peak
F_CONT_P1	f_{cont} of the first peak
F_CONT_P2	f_{cont} of the second peak
F_CONT_P3	f_{cont} of the third peak
DELTA_LFC	Difference in f_{cont} between best and 2nd best counter part