



Universiteit
Leiden
The Netherlands

Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection

Adam, R.; Vannier, M.; Maurogordato, S.; Biviano, A.; Adami, C.; Ascaso, B.; ... ; Zamorani, G.

Citation

Adam, R., Vannier, M., Maurogordato, S., Biviano, A., Adami, C., Ascaso, B., ... Zamorani, G. (2019). Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection. *Astronomy And Astrophysics (0004-6361)*, 627, A23. doi:10.1051/0004-6361/201935088

Version: Accepted Manuscript

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/84540>

Note: To cite this publication please use the final published version (if applicable).

Euclid preparation III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection

Euclid Collaboration, R. Adam^{1,2,3*}, M. Vannier², S. Maurogordato², A. Biviano⁴, C. Adami⁵, B. Ascaso⁶, F. Bellagamba^{7,8}, C. Benoist², A. Cappi^{2,8}, A. Díaz-Sánchez⁹, F. Durret¹⁰, S. Farrens¹¹, A.H. Gonzalez¹², A. Iovino¹³, R. Licitra^{6,14}, M. Maturi¹⁵, S. Mei^{6,14,16}, A. Merson^{16,17}, E. Munari^{4,18,19}, R. Pello²⁰, M. Ricci², P.F. Rocci², M. Roncarelli^{7,8}, F. Sarron¹⁰, Y. Amoura¹⁰, S. Andreon¹³, N. Apostolakis²¹, M. Arnaud^{11,22}, S. Bardelli⁸, J. Bartlett²³, C.M. Baugh²⁴, S. Borgani^{4,18,25}, M. Brodwin²⁶, F. Castander^{27,28}, G. Castignani^{6,29,2}, O. Cucciati⁸, G. De Lucia⁴, P. Dubath²¹, P. Fosalba^{27,28}, C. Giocoli^{7,8,30}, H. Hoekstra³¹, G. Mamon¹⁰, J.B. Melin¹¹, L. Moscardini^{7,8,30}, S. Paltani²¹, M. Radovich³², B. Sartoris¹⁸, M. Schultheis², M. Sereno^{8,7}, J. Weller^{33,34,35}, C. Burigana^{36,37,38}, C. S. Carvalho³⁹, L. Corcione⁴⁰, H. Kurki-Suonio⁴¹, P. B. Lilje⁴², G. Sirri³⁰, R. Toledo-Moreo⁴³, and G. Zamorani⁸

(Affiliations can be found after the references)

Received June 12, 2019 / Accepted –

Abstract

Galaxy cluster counts in bins of mass and redshift have been shown to be a competitive probe to test cosmological models. This method requires an efficient blind detection of clusters from surveys with a well-known selection function and robust mass estimates, which is particularly challenging at high redshift. The *Euclid* wide survey will cover 15000 deg² of the sky, avoiding contamination by light from our Galaxy and our Solar System in the optical and near-infrared bands, down to magnitude 24 in the *H*-band. The resulting data will make it possible to detect a large number of galaxy clusters spanning a wide-range of masses up to redshift ~ 2 and possibly higher. This paper presents the final results of the *Euclid* Cluster Finder Challenge (CFC), fourth in a series of similar challenges. The objective of these challenges was to select the cluster detection algorithms that best meet the requirements of the *Euclid* mission. The final CFC included six independent detection algorithms, based on different techniques, such as photometric redshift tomography, optimal filtering, hierarchical approach, wavelet and friend-of-friends algorithms. These algorithms were blindly applied to a mock galaxy catalog with representative *Euclid*-like properties. The relative performance of the algorithms was assessed by matching the resulting detections to known clusters in the simulations down to masses of $M_{200} \sim 10^{13.25} M_{\odot}$. Several matching procedures were tested, thus making it possible to estimate the associated systematic effects on completeness to $< 3\%$. All the tested algorithms are very competitive in terms of performance, with three of them reaching $> 80\%$ completeness for a mean purity of 80% down to masses of $10^{14} M_{\odot}$ and up to redshift $z = 2$. Based on these results, two algorithms were selected to be implemented in the *Euclid* pipeline, the Adaptive Matched Identifier of Clustered Objects (AMICO) code, based on matched filtering, and the PZWav code, based on an adaptive wavelet approach.

Key words. Cosmology: observations; large-scale structure of Universe – Galaxies: clusters: general

1. Introduction

Galaxy clusters are good tracers of the matter density peaks in the cosmic web. They additionally provide efficient tests for cosmological models as they form via gravitational collapse in the expanding Universe (for a review, see Allen et al. 2011). In particular, the number density of galaxy clusters as a function of mass and redshift enables us to constrain cosmological parameters primarily through the linear growth rate of perturbations. This has been proven to be very competitive and complementary to other probes (e.g., Vikhlinin et al. 2009; Rozo et al. 2010; Planck Collaboration et al. 2014; Böhringer et al. 2014; Mantz et al. 2015; Planck Collaboration et al. 2016c; de Haan et al. 2016). The spatial distribution of clusters can provide additional information to help constrain cosmological parameters via the measurement of the cluster-cluster two-point correlation function (e.g., Majumdar & Mohr 2004; Mana et al. 2013; Veropalumbo et al. 2014; Sridhar et al. 2017). In particular, clusters probe a redshift range that is sensitive to dark energy and hence they can be used to constrain extensions of the standard model. However, any cosmological inference using cluster

counts or spatial distribution requires accurate calibration of the halo mass function, an accurate knowledge of the cluster sample selection function, and primary observables that tightly correlate to cluster masses via scaling relations (including an understanding of the intrinsic scatter in the scaling relations). The calibration of the proper mass scale is also fundamental for cluster physics studies.

Galaxy clusters can be detected through their hot gas content, either from their X-ray emission (see e.g., Böhringer et al. 2001; Pacaud et al. 2016), or using their imprint in the Cosmic Microwave Background (CMB) via the thermal Sunyaev–Zel’dovich effect (tSZ, Sunyaev & Zel’dovich 1972) at millimeter wavelengths (e.g., Hasselfield et al. 2013; Bleem et al. 2015; Planck Collaboration et al. 2016a). In the optical (e.g., Kepner et al. 1999; Rykoff et al. 2014) or near-infrared (NIR; e.g., Eisenhardt et al. 2008; Wylezalek et al. 2013; Rettura et al. 2014) clusters can be identified using galaxy overdensities. Additionally, optical imaging and analysis methods have now reached the maturity to construct convergence maps via the weak lensing (WL) of background galaxies, where massive clusters appear as peaks (e.g., Gavazzi & Soucaill 2007; Shan et al. 2012; Jeffrey et al. 2018). In a cosmological context, the quest for a well-characterized cluster sample, preferably as complete

* Corresponding author: Rémi Adam

and as pure as possible, is important in quantifying the likelihood of cluster detections for a given set of cosmological parameters.

The properties of galaxy groups and clusters are also essential for understanding galaxy formation because they constitute the local environment in which a significant fraction of galaxies evolve (see, e.g., [De Lucia et al. 2012](#); [Raichoor & Andreon 2012](#)). Observations show that, at fixed stellar mass, cluster core galaxies present specific properties compared to field galaxies such as lower star formation rates, early-type morphologies and a tight red sequence up to redshift $z \sim 1$ (e.g., [Mei et al. 2009](#); [George et al. 2011](#); [Wetzel et al. 2013](#)). At higher redshifts, higher star formation rates are observed in cluster cores as well as more disturbed morphologies (e.g., [Brodwin et al. 2013](#); [Alberts et al. 2016](#); [Noirot et al. 2016](#)). A deeper understanding of the mechanisms that trigger such properties and their evolution will be achievable with future large-scale optical or NIR surveys such as *Euclid* ([Laureijs et al. 2011](#)), the Large Synoptic Survey Telescope (LSST, [LSST Science Collaboration et al. 2009](#)), the Javalambre-Physics of the Accelerated Universe Astrophysical Survey (J-PAS, [Benitez et al. 2014](#)), and the Wide Field Infrared Survey Telescope (*WFIRST*, [Spergel et al. 2015](#)), which will reach cluster masses down to a few $10^{14} M_{\odot}$ up to $z \sim 2$ ([Sartoris et al. 2016](#); [Ascaso et al. 2017](#)). Optical or NIR observation can also potentially select the most massive clusters at high redshifts (see e.g., [Andreon et al. 2009](#); [Brodwin et al. 2012](#)), and those are likely the place where the first massive galaxies form.

Euclid is a European Space Agency (ESA) mission planned for launch in 2021 that aims at providing a better understanding of the origin of the accelerated expansion of the Universe, particularly the nature of dark energy, dark matter, and gravity ([Laureijs et al. 2011](#); [Amendola et al. 2013](#)). Through its dedicated wide survey, *Euclid* will observe 15000 deg^2 , that is a large fraction of the sky (outside of the Galactic plane), in a wide optical band (VIS, down to magnitude 24.5 for a 10σ extended object) and three near-infrared bands (Y, J, H , down to magnitude 24 for a 5σ point-source). Deep surveys will cover about 40 deg^2 , which is two magnitudes deeper. Using the Near Infrared Spectrometer and Photometer (NIS) slitless spectrograph, photometric data will be complemented by spectroscopy, which is expected to release redshifts for several tens of millions of galaxies. Photometric redshifts that will be obtained by combination with ground based photometric surveys (such as the LSST, J-PAS or the Dark Energy Survey, DES, [Abbott et al. 2018](#)) will enable *Euclid* to detect galaxy clusters over a large range of masses and up to redshift ~ 2 . As an optical and NIR survey, the rest-frame optical richness of clusters will be the natural mass proxy, for which *Euclid* will be able to provide an internal calibration using WL mass estimates and velocity dispersion from spectroscopy using stacking techniques. A recent assessment of *Euclid* performance in terms of weak lensing mass estimates of ensemble clusters ([Köhlinger et al. 2015](#)) has shown that statistical uncertainties are expected to reach a very low level, and that usually predominant systematic errors such as multiplicative bias and additive bias are expected to be negligible. The richness estimates will also be complemented by other multiwavelength (X-ray, tSZ) mass proxies to reduce systematic uncertainties in the calibration. The combination of these properties should allow *Euclid* to push cluster cosmology to an unprecedented level (e.g., constraints of the order of a few percent on the dynamical evolution of dark energy or the growth factor parameter γ , [Sartoris et al. 2016](#)).

In order to reach these goals, several cluster finders have been developed within the *Euclid* consortium. It was then neces-

sary to develop a work frame to test and evaluate the performance of these different algorithms in the context of *Euclid*. Two main methodologies are generally used in the literature, both presenting advantages and limitations: 1) the use of end-to-end simulated data, aiming at matching the expected properties of the real data (e.g., [Koester et al. 2007](#); [Knobel et al. 2009](#); [Adami et al. 2010](#); [Old et al. 2015](#)), or 2) the injection of simulated clusters in a given existing data set (e.g., [Adami et al. 2000](#); [Goto et al. 2002](#); [Kim et al. 2002](#); [Rykoff et al. 2014](#); [Planck Collaboration et al. 2016a](#)). Given the rise of multiwavelength data-sets, the comparison of the cluster detections based on different tracers is also now a powerful way to cross-validate the selection functions (e.g., [Saro et al. 2015](#)). On one hand, the first method includes realistic projection effects associated with the spatial correlation between structures, while they are difficult to reproduce using the second method. This is particularly relevant in the case of cluster detection based on the galaxy distribution because the background is expected to be correlated with the targeted objects. On the other hand, the first method relies on the implementation of complex recipes to model the data, while the second method by construction is based on data. The second method is also more flexible regarding the modeling of the simulated cluster. Finally, arbitrary large volumes may in principle be created using the first method, while the second approach requires having in-hand data that are representative of the given survey under consideration, and large volumes to test the detection with sufficient statistics. Recently, the joint use of data and mocks has been shown to be extremely successful to fully account for correlated and uncorrelated background in the determination of richness ([Costanzi et al. 2019](#)), demonstrating the benefits of both approaches.

For the purpose of this paper, we use mocks to evaluate and compare the performance of cluster finders. This choice was motivated by several factors: i) mocks allow us to probe the whole redshift range that will be covered by *Euclid* on a wide-range of richnesses and masses; ii) they provide the distribution of halos of a given mass and redshift, which can be used as a truth table; and iii) they preserve the effect of the correlated background. We stress that the main limitation of this approach is the fact that simulations may not fully reproduce all the cluster properties, and the absolute performance derived may therefore be taken with caution. However, we found it the most operational way to compare the relative performance of the different algorithms on a common ground. The full methodology currently developed to determine the selection function and the related mass proxy will be addressed in future work.

The performance of the cluster finder algorithms has been tested and compared in a series of four Cluster Finder Challenges (CFC) between 2013 and 2017. The codes were tested on *Euclid* survey-like mock catalogs based on semianalytic models ([Merson et al. 2013](#); [Gonzalez-Perez et al. 2014](#)) and halo occupation modeling ([Carretero et al. 2015](#)). The positions of the mock clusters were unknown to the participants of the challenges. Through the years, the mock catalogs were refined to better represent the properties of galaxies within clusters. In particular, photometric redshifts were assigned to galaxies in order to run the codes as they would be run in the *Euclid* context. In the first challenges, photometric redshifts were assigned following a simplistic Gaussian distribution, while in the later ones, photometric redshift codes were used. At the end of the third cluster finder challenge, the methodology and analysis pipeline were sufficiently mature for a first assessment of the relative performance of the different codes. While eight cluster finder codes

in total were tested in the three preliminary challenges, only six of them took part in the final challenge described in this paper.

In this article, we present the methodology used to assess the performance of the codes and the results obtained from the final cluster finder challenge. The detection codes were applied blindly to a realistic galaxy mock, built using PhotReal (Ascaso et al. 2015) on the *Euclid* wide light-cone (Merson et al. 2013), which was considered to be the best compromise available in terms of angular size (300 deg^2), depth ($z > 2.5$), and realistic modeling of galaxy properties. We present the main assumptions and methodology of each of the competing codes and discuss the main properties of the simulated mock in the context of cluster detection. The code detections were matched to the true mock clusters and this information was used to evaluate the performance of the algorithms. Special care was given to the matching procedure by using several methods, allowing us to estimate the associated systematic uncertainties. In light of the mock properties, the performance comparison of the different algorithms participating in the challenge guided our selection of those now being validated and implemented in the *Euclid* pipeline. At this stage, we stress that the goal of this paper is not yet to compute a robust selection function and robust mass proxies, but instead, to compare the relative performance of different algorithms and to test different methodologies. The definition and assessment of the selection function and the best mass proxies will be addressed in future publications.

This paper is organized as follows. In Section 2, we present the competing algorithms. Section 3 describes the characterization of the simulations that are used. The matching procedure, of associating the detected clusters to the mock clusters, is detailed in Section 4, and the performance of the algorithms is given in Section 5. We discuss the results and the *Euclid* algorithm selection in Section 6. Conclusions are given in Section 7. A brief summary of the previous challenges, as well as the description of the previously employed codes are given in the Appendix. Throughout this paper, we assume a flat Λ CDM cosmology according to that used in the mock, with $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, and $\sigma_8 = 0.9$. All logarithmic quantities shown in this paper are defined using base 10. All the magnitudes in the paper are given in the AB system.

2. Galaxy cluster detection algorithms

The detection of galaxy clusters from photometric (or spectroscopic) surveys at optical and NIR wavelengths is a longstanding issue (see e.g., the pioneering work by Abell 1958). Several techniques have been developed, using different kinds of information. Some algorithms are based on the geometrical distribution of galaxies, both in projected coordinates and in photometric redshift space, while others also focus on known properties of cluster galaxies, such as colors, luminosities, and density profiles. Cluster finders are generally classified by methodology (or a combination of methodologies), of which a large variety exists in the literature. Some common examples include the use of the cluster red sequence (e.g., Gladders & Yee 2000; Rykoff et al. 2014), the presence of brightest cluster galaxies (BCG; e.g., Koester et al. 2007), percolation algorithms (e.g., Dalton et al. 1997), matched filtering (e.g., Postman et al. 1996; Olsen et al. 2007), Voronoi tessellation methods (e.g., Ramella et al. 2001), friends-of-friends (FoF; e.g., Wen et al. 2012), the use of smoothing kernel techniques (e.g., Gal et al. 2003; Mazure et al. 2007), or wavelet filtering techniques (see e.g., the pioneering work of Eisenhardt et al. 2008). These techniques have been ex-

tensively used to build large samples of clusters (e.g., Gilbank et al. 2011) and have also led to the discovery of some massive clusters at high redshifts (e.g., Stanford et al. 2012). All detection techniques present advantages and drawbacks regarding selection effects, however different techniques are often complementary to one another. For instance, searching for the presence of a red sequence can be an efficient way to detect clusters at low and intermediate redshifts. This property, however, is expected to fade at higher redshifts (e.g., Strazzullo et al. 2016, and references therein) making it less effective for detecting distant clusters. For a review on cluster detection, see for example Gal (2006), or for a detailed discussion about the necessary features of galaxy cluster finders in the context of large photometric surveys, see for example Rykoff et al. (2014).

The detection of galaxy clusters in the *Euclid* survey will be largely driven by photometric data. Indeed, analytical estimates (Sartoris et al. 2016) have shown that the mass detection limits obtained using spectroscopic redshifts are significantly higher than those obtained with photometry. Spectroscopic redshifts may also be used to improve the detection procedure, nevertheless this has not been taken into consideration for this work and is left for future studies. Spectroscopic data will, however, be used to confirm and refine the redshifts of the clusters detected by photometry.

Six algorithms participated in the final CFC. They were all blindly applied to a simulated mock catalog (see Section 3) to provide a cluster catalog with the coordinates of the objects (sky coordinates: right ascension, RA, and declination, Dec, and redshift), a mass proxy (typically the richness) and a ranking of the likeliest true detections (mainly by signal-to-noise ratio, S/N). Four algorithms also provided the probability of the cluster member galaxies associated with each detected cluster. The names of the cluster finders, as used hereafter, and their main detection principles are provided in Table 1. The following subsections provide an overview of the methodology and the assumptions used by each code.

2.1. AMASCFI: Adami, Mazure & Sarron cluster finder

The Adami, Mazure & Sarron cluster finder (AMASCFI) algorithm (Sarron et al. 2018) searches for clusters in large multi-band imaging surveys using photometric redshift (z_{phot}) tomography. As an input, the AMASCFI algorithm requires a galaxy catalog with sky positions (RA, Dec) and photometric redshifts. The photometric redshift catalog is first divided in redshift slices of variable width according to the evolution of the photometric redshift error, $\sigma_{z_{\text{phot}}}(z_{\text{spec}})$, which is estimated using spectroscopic redshifts from the calibration field (see Section 3). All slices overlap by 0.05 in redshift, taken as a constant so that the cluster photometric redshifts are sampled with the same resolution whatever the redshift. Galaxy density maps are built for each redshift slice, based on an adaptive kernel technique, with an initial kernel size (diameter) fixed at 1.5 Mpc. This way the adaptive kernel size in the densest region (corresponding to galaxy clusters) is about 1 Mpc (i.e., the typical size of cluster cores). Structures in these density maps are detected using the source extraction software, SExtractor (Bertin & Arnouts 1996), in the different redshift bins with a detection threshold set to a given number of galaxies per Mpc^2 . The initial structures are then assembled into larger structures using a minimal spanning tree FoF algorithm (see Adami & Mazure 1999). Any two detections less than 1 Mpc apart and with $\Delta z \leq 0.05$ are merged. A detailed description of each step of the algorithm, as well as a discussion of

Table 1: Summary of properties and names of eight cluster finder algorithms that participated in CFC. The properties listed here correspond to those of the final CFC. All algorithms performed redshift slicing or made use of a grid, and all rely on the *H*-band in the case of the final CFC. RedGOLD and Voronoi did not participate in the last challenge for reasons not related to their performance in the earlier ones.

Name	CFC participation	Detection principle	Main reference	Cluster properties assumptions	Use of calibration field	Membership
AMASCFI	1,2,3,4	Adaptive kernel	Adami & Mazure (1999)	Typical size and m_H^* calibration	Yes	X
AMICO	1,2,3,4	Optimal filtering	Bellagamba et al. (2018)	LF and profile	No	✓
HCFA	3,4	Hierarchical finder	Díaz-Sánchez (in prep.)	Typical size only	No	✓
PZWav	1,2,3,4	Wavelet adaptive	Gonzalez (2014)	Typical size and m_H^* evolution	No	X
sFoF	1,2,3,4	Friends-of-friends	Farrens et al. (2011)	None	Yes	✓
WaZP	1,2,3,4	Wavelet	Benoist (2014)	Typical size and m_H^* evolution	Yes	✓
RedGOLD	1,2	Red sequence	Licitra et al. (2016a)	–	–	–
Voronoi	1	Voronoi tessellation	Iovino (in prep.)	–	–	–

the influence of the choice of parameters can be found in Sarron et al. (2018).

The sky coordinates (RA, Dec) and redshift of each candidate cluster are taken to be the mean of each of its individual merged detections weighted by its galaxy number density. For each redshift slice, the S/N of detected peaks is computed from the 2D density map as $(\langle n_{\text{cluster}} \rangle A - \langle n_{\text{field}} \rangle A) / \sqrt{\langle n_{\text{field}} \rangle A}$, where $\langle n_{\text{cluster}} \rangle$ and $\langle n_{\text{field}} \rangle$ correspond to the average number density of galaxies per Mpc^2 in a slice of width Δz for cluster and field area, respectively, and A is the cluster area (taken to 500 kpc radius) projected on the sky. For each cluster candidate, the final S/N is taken as the maximum S/N of its individual merged detections. The richness λ_{det} is computed from a modified version of the Licitra et al. (2016a) estimator. AMASCFI first counts the number of galaxies with $m_H < m_H^* + 2.5$ in a cylinder of radius $R_{\text{det}} = 1 \text{ Mpc } h^{-1}$ and length $\pm 2\sigma_{z_{\text{phot}}}$ around the cluster center, and removes the galaxy background contribution. The knee magnitude of the luminosity function (LF), m_H^* , was calibrated using the value measured for the Coma Cluster obtained by de Propriis et al. (1998). It then iteratively rescales the detection radius as $R_{\text{det}} = (\lambda_{\text{det}} (< R_{\text{det}}) / 100)^{0.2}$ until convergence. For the last CFC, the rank was determined by sorting the S/N values. The richness was used to establish the relative rank for objects with identical S/N values. AMASCFI was applied to the CFHTLS in Sarron et al. (2018) and the previous version of the AMASCFI algorithm (AMACFI) was used to search for clusters in the CFHTLS (Mazure et al. 2007; Adami et al. 2010; Durret et al. 2011) and in the SDSS Stripe 82 data (Durret et al. 2015).

2.2. AMICO: Adaptive Matched Identifier of Clustered Objects

The Adaptive Matched Identifier of Clustered Objects (AMICO) algorithm (Bellagamba et al. 2011, 2018) is an enhanced matched filter algorithm that looks for cluster candidates by convolving the 3D galaxy distribution with a redshift-dependent filter. The input of the algorithm is a galaxy catalog that includes sky coordinates (RA, Dec), photometric redshifts and magnitudes. The filter is defined on the basis of a cluster and noise model that has the purpose of amplifying the contrast between the two components. Originally this filtering method was used to detect galaxy clusters in weak lensing data (Maturi et al. 2005). The noise is modeled by assuming a spatially uniform LF, while the cluster model is the combination of a cluster galaxy LF and a galaxy density profile. In the CFC, AMICO considered only the *H*-band for detection, but it can use any other magnitude or a combination of two or more. It also accounts for the full shape of the photometric redshift probability distribution function (PDF), $P(z)$, provided by the mock. The convolution of the

galaxy distribution with the AMICO filter generates a 3D amplitude map, whose peaks represent the detections. In addition to standard matched filter algorithms, AMICO defines a membership probability for each galaxy to belong to a given detection. It uses this information to remove signals in the original amplitude map in order to search for further detections, which might be blended with other structures, without any further assumptions. This has proven to be an efficient method to disentangle close-by objects.

The output sky coordinates (RA, Dec) and redshift of the candidate clusters are given by the position of the peaks in the likelihood on the 3D grid. The uncertainty on the amplitude is derived from the expected variance in the measurement, due to the background fluctuations and the shot-noise in the cluster galaxy distribution. The S/N associated to the candidate clusters is then the ratio of the amplitude over its uncertainty. The mass proxy provided by AMICO is the amplitude, a measure of the cluster galaxy abundance in units of the cluster model. Detections are ranked according to their S/N. We note that AMICO can provide another mass proxy, given by the sum of the membership probabilities for each detection (a measurement of the richness, see Bellagamba et al. 2019), but this quantity was not used in this work. AMICO was recently used to identify galaxy clusters in the Kilo Degree Survey (KiDS, Radovich et al. 2017; Maturi et al. 2019).

2.3. HCFA: Hierarchical Cluster Finder Algorithm

The Hierarchical Cluster Finder Algorithm (HCFA) algorithm (Díaz-Sánchez, in prep.) searches for overdensities of galaxies using different angular scales in a hierarchical approach. The HCFA algorithm requires only the position and the photometric redshift of the galaxies as inputs. It first uses overlapping redshift bins of size $\Delta z = 0.05$ (as for AMASCFI) to identify the galaxies that are in local overdensity regions. Each galaxy is then labeled with its local density, n_g , according to the galaxies in its neighborhood. HCFA uses a primary angular scale of 0.2 Mpc for this purpose. A critical density n_{gc} is defined as $3\sigma_{n_g}$ above the mean local density $\langle n_g \rangle$, $n_{gc} = 3\sigma_{n_g} + \langle n_g \rangle$, where σ_{n_g} is the standard deviation of the local galaxy density field, and galaxies labeled with lower densities are removed from the sample. The remaining galaxies are merged using a FoF algorithm with an angular linking scale equal to the primary one. The overdensity factor is calculated for each resulting group of galaxies. Membership probabilities are defined according to the local density of each galaxy. Groups with densities lower than the critical density are removed and the FoF algorithm is repeated iteratively, increasing the angular linking scale, until groups do not merge any more or

the linking scale reaches 0.6 Mpc. In this way, HCFA identifies galaxy clusters composed of hierarchical overdensities. The algorithm uses a sky tiling of 36 arcmin² (chosen for convenience) and tiles are processed in parallel.

The cluster candidate centroids are calculated taking into account all the galaxies in the cluster, while the redshift is given by the mean redshift of the galaxies. A S/N is defined for each galaxy as $(n_g - \langle n_g \rangle) / \sigma_{n_g}$. From this definition, the S/N of the candidate clusters are set to the mean S/N of the five galaxies with the highest S/N values in the cluster. A minimum of five galaxies are required in order to define a candidate cluster. The richness is given by the total over-density factor of the cluster, i.e., the number of galaxies in the cluster multiplied by the S/N of each galaxy. The candidate clusters are ranked according to the S/N. The HCFA algorithm has not yet been applied to real data.

2.4. PZWav

The cluster finding algorithm PZWav (Gonzalez 2014) is a wavelet-style algorithm that searches for overdensities on fixed physical scales. PZWav requires a galaxy catalog with sky coordinates, photometric redshifts, and magnitudes. It uses a difference-of-Gaussian smoothing kernel and incorporates for each galaxy the full probability distribution associated with the photometric redshift, $P(z)$. As a preprocessing step, the galaxy catalog is culled to contain only galaxies brighter than a given limit, taken as $m_H < m_H^* + 2$ in *H*-band, so that galaxies out to $z = 1.5$ are selected down to the same limit, as traced by any model of galaxy evolution. This preprocessing step minimizes the redshift dependence of the mass threshold for cluster detection. After this preprocessing is complete, the algorithm first constructs a series of redshift slices spanning the redshift range of interest, and then inserts each galaxy into these redshift slices, weighted by the probability that the galaxy lies at a given redshift. These density maps are next convolved with a difference-of-Gaussians smoothing kernel of a fixed physical size, which is approximately matched to the physical size of cluster cores. A second set of density maps is also constructed for which the redshift probability distributions have been randomly shuffled relative to the positional information. These random density maps are used for bootstrap simulations to calculate a uniform noise threshold as a function of redshift that is independent of the mean galaxy density. Galaxy cluster candidates are next identified in each redshift slice, and these detections are merged across the redshift slices. All detections that lie near the edge of the survey field are rejected, and redshift estimates are refined for each cluster using a secondary code that sums the probability distributions of all galaxies within a fixed radius of the cluster detection.

The cluster centroids come directly from the smoothed density maps, corresponding to the peak location of each detected overdensity. Cluster redshifts are derived by computing the σ -clipped median photometric redshift from all galaxies that lie within 30'' of the centroid and lie within $\Delta z = 0.12$ of the redshift slice in which a cluster is detected. The direct observable from this search is the peak amplitude of each detected overdensity, which can be taken as a proxy for richness. Candidates are ranked by this peak amplitude. The version of PZWav used for the challenges did not calculate the S/N, reporting only the peak amplitude. The current version of the code calculates the S/N based upon the fluctuations in the random maps. This algorithm is based upon the approach initially developed for the

IRAC Shallow Cluster Survey (Elston et al. 2006; Eisenhardt et al. 2008), also used in the work of Stanford et al. (2012), but has been optimized and refined to work efficiently with *Euclid*-like data.

2.5. sFoF: Friends-of-friends

The sFoF algorithm is a friends-of-friends galaxy cluster detection algorithm (Farrens et al. 2011) that follows the principles established by Huchra & Geller (1982) and later modifications implemented by Botzler et al. (2004). The algorithm operates using an input galaxy catalog with either spectroscopic redshifts (3D: using sky coordinates and redshifts) or photometric redshifts (2+1D, as in the present case), using sky coordinates stacked in bins of photometric redshift. All of the internal operations are performed in angular space and no assumptions are made about the nature of clusters of galaxies (e.g., size, color, shape). Two primary free parameters, the transverse linking and the line-of-sight linking lengths, determine the total number of cluster candidates and their corresponding properties. These linking parameters change as a function of redshift to account for selection effects, which in turn provides a redshift independent richness estimate for each cluster candidate. The parameters were optimized using the calibration field provided with the mock (see Section 3). Each FoF group galaxy is marked as a cluster member and its membership probability is set to unity, while non cluster members have a membership probability that is set to zero. The code implements k-dimensional tree and Open Multi-Processing routines to improve the performance of a single run.

The cluster candidate coordinates (RA, Dec and redshift) are obtained from the median of the member positions. The S/N is computed as $(\lambda_{\text{det}} - A n_{\text{field}}) / \sqrt{A n_{\text{field}}}$, where λ_{det} is the estimated richness, A is the cluster area projected on the sky, and n_{field} is the galaxy background level at the cluster redshift. The richness is given by the number of FoF objects found for a given cluster, which is also the sum of the membership probabilities. Because the linking parameters change as a function of redshift, this roughly gives a redshift independent estimate. Candidate clusters were ranked according to the richness. The sFoF algorithm was applied to the 2SLAQ spectroscopic survey (Cannon et al. 2006) of potential luminous red galaxies in Farrens et al. (2011).

2.6. WaZP: Wavelet Z-Photometric cluster finder

The Wavelet Z-Photometric cluster finder (WaZP) algorithm (Benoist 2014; Dietrich et al. 2014) is an optical cluster finder based on the identification of galaxy overdensities in (RA, Dec, z_{phot}) space. WaZP requires a galaxy catalog with sky coordinates (RA, Dec), photometric redshifts and magnitudes. The detection process makes no assumptions on the LF of cluster galaxies nor on the galaxy density profile. From an operational point of view the WaZP algorithm goes through the sequence described below. The galaxy catalog is sliced along the photometric redshift axis in overlapping redshift bins of variable widths controlled by the scatter of $P(z)$. In each slice, galaxies are weighted by the fraction of their PDF intersecting the slice. In addition, in the context of this work, detection was performed only using galaxies with $m_H \leq m_H^* + 1$. The resulting projected galaxy distribution is then pixelated on a grid with a physical step size of 1/16th of a Mpc. The pixelated galaxy catalog is filtered using the wavelet task MR_FILTER from the multiresolution package MR/1 (Starck et al. 1998). This task incorporates

a statistically rigorous treatment of the Poisson noise, which makes it possible to keep significant structures in an appropriate scale range. Here structures with scales up to 1 Mpc are selected and a 3σ iterative multiresolution thresholding with a B-spline wavelet transform is applied. From each wavelet map, peaks are extracted and merged with peaks from consecutive slices to produce a final cluster list.

Each peak detected in the projected filtered maps is characterized by i) a position defined as the mode of the peak, ii) a radius R_{det} defined as the mean extent of the peak, iii) a redshift defined as the median redshift of the photometric redshifts selected within a projected distance $\leq R_{\text{det}}$ from the center and within $\pm 3\sigma_{z_{\text{phot}}}$ around the mean redshift of the map, and iv) a S/N defined as $(n - \langle n \rangle) / \sigma_{\text{bg}}$ where n and $\langle n \rangle$ are the galaxy density within 300 kpc from the peak center and the galaxy local background density respectively. The quantity σ_{bg} is given by the second order moments of galaxy counts in cells. When a cluster is detected in several consecutive slices, it is associated to the peak with the largest S/N. For each cluster, membership probabilities are computed following the prescription given in [Castignani & Benoist \(2016\)](#), based here on a local background density modeling. Membership probabilities are computed up to a radius corresponding to a given galaxy density contrast. Finally each cluster is characterized by a richness defined as the sum of the membership probabilities for galaxies with a magnitude $m_H \leq m_H^* + 1$. Clusters are ranked according to their S/N. The WaZP algorithm was applied to N-body simulations in [Dietrich et al. \(2014\)](#) and to the CFHTLS data to search for optical counterparts to the XXL survey ([Pierre et al. 2016](#)) X-ray clusters ([Benoist et al., in prep.](#)).

3. *Euclid* mock galaxy catalog

The final *Euclid* CFC made use of a main mock galaxy catalog ([Ascaso et al. 2015](#)) in order to test the behavior of the detection algorithms on *Euclid*-like data. This mock includes photometric redshifts, z_{phot} , and their errors. It was limited to *H*-band magnitudes brighter than $H_{\text{AB}} = 24$ to mimic the context of the *Euclid* wide survey ($H_{\text{AB}} = 24$ for 5σ point-source). A 20 deg^2 region including both photometric and spectroscopic redshifts was also provided as a calibration field for the photometric redshifts or for the detection code parameters. While it is not the purpose of this paper to make an assessment of the validity of the semianalytic models on which the mock is based, we do aim to verify the reliability of the model predictions. This is done in order to quantify how realistic the performance of the cluster finders are when applied to the mock. We discuss the construction of the mock in Section 3.1.

3.1. Construction of the mock galaxy catalogs

We placed some constraints on the properties of the mock as we aimed to test the performance of the cluster finders at high redshift (up to about 2) and high mass (larger than about $10^{14} M_{\odot}$) in the *Euclid* regime. In order to satisfy these requirements, the mock has to be complete in magnitude to at least $H_{\text{AB}} = 24$, to cover a redshift range up to $z \gtrsim 2$, and to have a reasonable sky coverage in order to get enough statistics on the high mass and high redshift clusters. We therefore chose a parent sample of 500 deg^2 from which we extracted a 300 deg^2 mock. Finally, this mock was blinded by applying a rotation and translation.

3.1.1. Galaxy catalog

The galaxy catalog was extracted from the [Ascaso et al. \(2015\)](#) mock, which was based on the *H*-band wide light-cone from [Merson et al. \(2013\)](#). The light-cone was generated from the Millennium simulation ([Springel et al. 2005](#)) using semianalytical modeling of galaxy formation with the GALFORM model ([Lagos et al. 2012](#)). The mock was reprocessed with the software PhotReal ([Ascaso et al. 2015](#)) to obtain realistic galaxy photometry compliant with *Euclid* depth in *YJH* (down to magnitude 24 at 5σ , point sources) and *grizY* (down to magnitudes 25.2, 24.8, 24.0, 23.4 and 21.7 at 10σ , extended sources), assuming complementary ground-based DES data ([Mohr et al. 2012](#)). This corresponds to the pessimistic case in [Ascaso et al. \(2015\)](#), as opposed to the combination of the *Euclid* observations with deeper ground-based photometry from LSST (the optimistic case in [Ascaso et al. 2015](#)). In this sense the performance derived hereafter is expected to be conservative.

The photometry was also modified by PhotReal using a set of empirical templates to fit observed spectral distributions and make the galaxy colors, luminosity and mass functions more consistent with current observations (see [Ascaso et al. 2015](#), for more details). Photometric redshifts were estimated using the Bayesian Photometric Redshifts software (BPZ, [Benítez 2000](#); [Benítez et al. 2004](#); [Coe et al. 2006](#)) applied to the PhotReal photometry. The most likely redshifts (PDF peaks) were derived, as well as their probability distribution functions.

We note that the magnitude cut applied to the mock used in the present paper introduces an extra idealization. Indeed, in practice the *Euclid* catalog will extend to fainter magnitudes (albeit being incomplete), which may benefit to the detection codes, in particular for the detection of high redshift clusters. In this sense, the results presented in this paper are conservative in terms of performance, as the magnitude cut applied limits the sampling of the luminosity function at high redshift (however still reaching $m^* + 1.5$ at redshift 2). In addition, accurate photometry in crowded cluster fields, with the intra cluster light also contributing to the background, is a real challenge as shown in recent studies based on Hubble Space Telescope observations (e.g., [Molino et al. 2017](#)). Such effects, which are not included in the mock used in this paper, may boost the photometric redshifts uncertainties of the corresponding galaxies, and we leave their detailed investigation for future work, when the end-to-end *Euclid* simulations including all observational effects, the final pattern of ground-based complementary observations, and the estimation of photometric redshifts performed with the *Euclid* code, will be available.

3.1.2. Mock cluster catalogs

Dark matter halos were identified in the simulation using the algorithm defined in [Jiang et al. \(2014\)](#), such that galaxies were given a group identifier and the central galaxies were marked. A cluster catalog was thus constructed by grouping galaxies that belonged to the same halo, using their unique identifiers. The coordinates of each cluster were taken to be those of the central galaxy, both in sky coordinates and redshift. We also observed that defining the mock cluster center using the barycenter of the member galaxies marginally impacts the results presented in this paper and differences are discussed hereafter whenever relevant. For each mock cluster we calculated the quantities RA_{min} , RA_{max} , Dec_{min} , Dec_{max} , i.e., the minimum and maximum right ascension and declination of the members. This defines a rect-

angular area that includes all the galaxies belonging to a given mock cluster.

The mock cluster masses, *Dhalo* (M_{DH}), were also defined according to Jiang et al. (2014). The M_{DH} values are related to the masses that are generally used in observations, such as M_{200}^1 . The median ratio between M_{DH} and M_{200} is equal to about 1.25 and the distribution remains confined between $\gtrsim 1$ and $\lesssim 1.5$ at 90% C.L., being fairly flat (Jiang et al. 2014). We note that in Jiang et al. (2014), the mass ratio is well characterized up to $M_{\text{DH}} \simeq 10^{14} h^{-1} M_{\odot}$. Given the smooth evolution of the ratio with mass over several orders of magnitude, we assume that extrapolation is accurate up to the high mass tail considered here, $M \sim 10^{15.5} M_{\odot}$. The final mock cluster catalogs were constructed by selecting all clusters down to masses of $10^{13.25} M_{\odot}$. The implications of this limit on our results is further discussed in sections 5 and 6. Hereafter, the masses are referred as M .

The characteristic radius was estimated as $\tilde{R}_{200} \equiv [M / (\frac{4}{3}\pi 200 \rho_c)]^{1/3}$. This quantity is related to the mass of each mock cluster and uses the critical density at the cluster redshift, ρ_c , as computed from the mock cosmological parameters in the flat Λ CDM model. Because the masses we used are not defined as M_{200} , our estimates of R_{200} are biased high by around 8% for the median of the cluster population, and remain less than 17% larger at 95% C.L. It should be noted, however, that these $\tilde{R}_{200}$ values were only used to associate detected clusters to mock clusters and hence this does not significantly affect our results, as discussed further in Section 4.

3.2. Properties of galaxies and galaxy clusters in the mocks

To facilitate the interpretation of the results of the final CFC and to validate the simulations for our purposes, we explore the properties of the mock in terms of photometric redshift reconstruction, mass-richness relation, cluster galaxy density profiles and galaxy cluster LF. An analysis of the galaxy properties in the mock is provided in Ascaso et al. (2015). In the following subsections we complement this analysis, particularly with regards to cluster environment.

3.2.1. Photometric redshift properties

The precision of the photometric redshift estimates is expected to have a significant impact on cluster finder performance. Clusters appear as overdensities not only in projected space, but also in redshift space, information that is used by the detection algorithms via the photometric redshifts. Ascaso et al. (2015) validated BPZ photometric redshifts comparing them to spectroscopic redshifts and assessing their performance in terms of resolution and outliers (see Section 5 of their paper and Tables 1 and 2). We briefly summarize their results and present an internal validation performed in the context of the CFC. Ascaso et al. (2015) showed that for the *Euclid pessimistic* case $\sigma_{\text{NMAD}} \leq 0.03$ for galaxy $m_H < 22.5$ and increases up to $\sigma_{\text{NMAD}} \sim 0.08$ at $m_H \sim 24$, using the normalized median absolute deviation (NMAD)². When considering all magnitudes up to $m_H = 24$, $\sigma_{\text{NMAD}} \leq 0.045$ for redshift $z < 1.5$ and $\sigma_{\text{NMAD}} \sim 0.06$ at $1.5 < z < 3$. These limits increase when using the *odds* parameter in BPZ (not used in the CFC). In terms of outliers, the

Euclid pessimistic case shows a rate of outliers in the range 10–20%, with the highest fractions in the redshift ranges $0.5 < z < 1$ and $2 < z < 3$. These results are shown in Ascaso et al. (2015) Tables 1 and 2 as a function of galaxy magnitude and redshift, and in Figures 17 to 22. As a general comment, the photometric redshift resolution of the *Euclid optimistic* case is a factor of two to five better than the pessimistic case both in terms of photometric redshift accuracy and bias.

We hereafter present the internal challenge validation of the photometric redshift quality in the simulation. For this, we follow Ricci et al. (2018), adapted from Ilbert et al. (2006)³. For each redshift bin, we compute the difference $z_{\text{phot}} - z_{\text{true}}$, and use the resulting distributions to extract the bias, the catastrophic failure fraction and the dispersion. Here, z_{true} refers to the true spectroscopic redshifts. These values account for peculiar velocities, which are known for all the galaxies in the simulation and are not affected by selection effects. The bias is computed as the median of the distribution. The outlier fraction is given by the fraction of objects satisfying $|z_{\text{phot}} - z_{\text{true}} - \text{bias}| > 0.15(1 + z_{\text{true}})$. The dispersion is computed both using NMAD as in Ascaso et al. (2015), and percentiles by integrating the distributions up to a 68.2% confidence level on the positive and negative parts. We also reproduce this analysis after removing galaxies with H -band $m_H > 23$ magnitude to highlight the effects of contamination from low S/N objects. We note that below this limit, the distribution remains fairly stable. Similarly, we reproduce this analysis by selecting cluster member galaxies above a given halo mass, to investigate potential environmental effects.

Figure 1 shows the comparison between the true spectroscopic redshift, z_{true} , and the photometric redshifts z_{phot} , for a randomly selected subsample of galaxies from the mock ($\sim 10^5$ galaxies are shown). This figure also provides the bias and the two estimates of the dispersion. Figure 2 shows the redshift evolution of the catastrophic outlier fraction (top panel), the bias (central panel) and the different estimates of the dispersion (bottom panel) for the full mock and after removing objects with $m_H > 23$. The left panel includes cluster and field galaxies while the right panel focuses on cluster member galaxies, belonging to haloes of mass larger than $10^{14} M_{\odot}$. We measure the overall mean photometric uncertainty to $\sigma_{z_{\text{phot}}} = 0.050(1 + z_{\text{true}})$. The dispersion increases by a factor of ~ 2 and becomes very asymmetric at $z_{\text{true}} \sim 0.5 - 0.6$. It also increases by a similar amount at redshifts below 0.2 and above 2.5 for the full catalog, but remains relatively flat for the high S/N catalog ($m_H < 23$). The bias becomes large where the photometric uncertainties are large, even for the $m_H < 23$ catalog. The fraction of catastrophic redshifts is small at redshifts above 0.8 ($\lesssim 0.05$ even for the full catalog, and about 0.01 for the high S/N catalog). However, it becomes large at lower redshifts, reaching up to 20% for the full catalog and 15% for the $m_H < 23$ catalog. The distribution remains very similar in the case where cluster member galaxies are selected, independently of the exact value adopted for the mass cut. We note that the overall quality of the photometric redshifts measured corresponds to the pessimistic case, as expected from the catalog used. In the context of *Euclid*, the standard deviation of the photometric redshifts with respect to the true redshifts is required to be $\sigma_z/(1 + z) < 0.05$, keeping as a goal $\sigma_z/(1 + z) < 0.03$ (Laureijs et al. 2011). Similarly, the catastrophic failures requirement is less than 10% beyond

¹ The mass M_{200} corresponds to the mass enclosed within a radius R_{200} , within which the mean density of the cluster is equal to 200 times the critical density of the Universe at the cluster redshift.

² The NMAD associated to the variable X is defined as $\sigma_{\text{NMAD}}(X) = 1.48 \text{ median}|X - \text{median}(X)|$.

³ See also the photometric redshift release explanatory document http://cesam.lam.fr/cfhtls-zphotos/files/cfhtls_wide_T007_v1.2_Oct2012.pdf

$0.15(1+z_{\text{true}})$, while the goal is to keep this less than 5% beyond $0.15(1+z_{\text{true}})$. Our internal validation is consistent with the mock validation performed in [Ascaso et al. \(2015\)](#) where a more optimistic case is also presented in addition to the pessimistic one used here. We note that the large number of outliers, the large bias and the large dispersion at redshifts below 0.3, above 2.3 or near 0.6 are largely due to the fact that no u -band is used in the pessimistic case, while it would be available in the optimistic case.

Based on the photometric redshift properties of the catalog, we expect cluster finder detection properties to be altered in the redshift range in which the catastrophic outlier fraction is large ($z_{\text{true}} \sim 0.5 - 0.6$, and $z_{\text{true}} \lesssim 0.2$). This is even more true for clusters with fewer member galaxies (i.e., at lower masses). This alteration might show up as an increased number of false detections or larger uncertainties in the redshift recovery of the clusters, depending on how the photometric redshifts are used by the finders. The bias can also affect the matching performed to associate the detections to the true clusters (see Section 4). At redshifts $0.8 < z_{\text{true}} < 2$, the photometric redshift distribution is nearly Gaussian (with small bias and a small catastrophic outlier fraction). Therefore, the cluster finders are expected to behave well despite the fact that the larger photometric errors and the lower number of galaxies, reduced by redshift dimming, should impact the completeness.

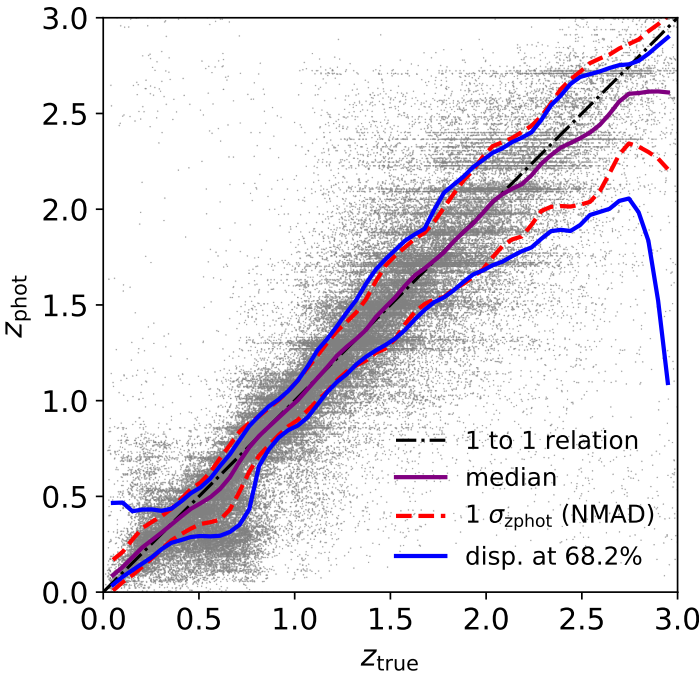


Figure 1: Comparison between photometric redshift, z_{phot} , and true spectroscopic redshifts, z_{true} . The bias is shown by the purple solid line, the NMAD is shown as the red dashed line, and the dispersion computed as percentiles is shown by the blue solid line. The black dashed-dotted line provides the one-to-one relation for reference.

3.2.2. Mass-richness relation

The richness of galaxy clusters is a fundamental quantity derived from optical or NIR surveys. It generally serves as the primary mass proxy and its normalization is tightly related to the detection performance at a given mass. In the context of the CFC, it

was necessary to characterize the mass-richness relation of the mock itself in order to estimate the scatter introduced to richness measurements (see Section 5). See also the work by [Ascaso et al. \(2017\)](#) for the characterization of the cluster total stellar mass as a cluster mass proxy, using the same mock.

For each mock cluster, we compute an estimate of the richness as the number of galaxies associated to the halo as

$$\lambda_{\text{mock}} = N_{\text{gal}}(m_H < m_{H, \text{ref}}^*(z_{\text{true}}) + 2). \quad (1)$$

In order to account for a redshift dependence of the richness definition, through the magnitude evolution, we exclude galaxies with m_H larger than $m_{H, \text{ref}}^* + 2$. This allows us to have a complete sample up to $m_H = 24$ at redshift 2.5 (see also the discussion on the LF in Section 3.2.4). The reference magnitude $m_{H, \text{ref}}^*$ is derived from the passive evolution of a starburst galaxy with a formation redshift $z_{\text{form}} = 3$ taken from the PEGASE2 library (burst_sc86_zo.sed, [Fioc & Rocca-Volmerange 1997](#)). It is calibrated using the value of K^* at redshift 0.25 derived by [Lin et al. \(2006\)](#) from an observed cluster sample. The validity of this evolution is addressed in Section 3.2.4 (see also Figure 5, right panel) and the exact $m_{H, \text{ref}}^*$ model used to compute λ_{mock} has a negligible impact on our results, especially given that it reproduces well the trend seen in the mock at the relevant redshifts.

In Figure 3, we provide an example of the scaling between the mass and the richness, computed for all clusters in the redshift range $[0.5 - 0.75]$. The mass-richness relation is modeled by power law and fitted using the bivariate correlated errors and intrinsic scatter (BCES, [Akritas & Bershady 1996](#)) method. The best-fit model is subtracted from the data and the residual is used to compute the scatter in the richness at fixed mass. The blue and purple dots provide the median and mean richness of the corresponding mass bin, while the error bars represent the scatter computed as the NMAD and the standard deviation, respectively. While the standard deviation is accurate for lognormal scatter, the NMAD is more robust to outliers and we use it as the baseline. The differences between the two methods are insignificant. The slope is consistent with unity within a few percent at all redshifts. The scatter does not significantly evolve with redshift (not shown), but it does decrease linearly with $\log M$ ($\sigma_{\log \lambda} \simeq 0.1$ at $M = 10^{13.5} M_{\odot}$ and $\sigma_{\log \lambda} \simeq 0.05$ at $M = 10^{14.5} M_{\odot}$). This intrinsic scatter will be later used when quantifying the scatter introduced by the detection algorithms in Section 5. We observe outliers at low richness in the scaling relation when using the mock cluster catalog based on the barycenter of cluster galaxies (not shown). They correspond to clusters that are on the edge of the footprints since their number of member galaxies is generally truncated while their mass remains the same. In principle, these clusters also affect the detections, but we have observed that they have a negligible impact on the global performance presented in this paper. In practice, the *Euclid* survey will be affected by masks, or varying depth, but at this stage not all the algorithms are able to handle such effects and we leave the investigation of their impact on the detection of galaxy clusters for future work.

3.2.3. Cluster galaxy density profile

The radial structure of galaxy clusters is a key property, which may affect any cluster finder. Since the CFC detection algorithms are driven by photometric data with photometric redshift uncertainties that are much larger than the cluster extent along the line of sight, they are mainly sensitive to the projected galaxy radial

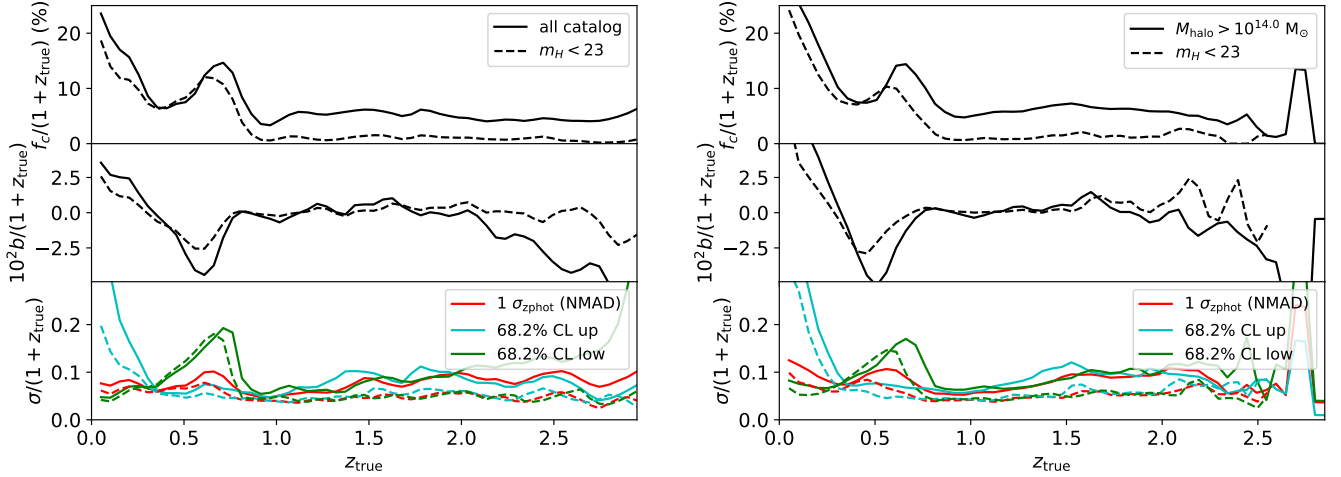


Figure 2: Redshift evolution of the catastrophic outlier fraction (f_c , upper panel), the bias (b , middle panel), and different estimates of the dispersions (σ , lower panel) as a function of spectroscopic redshift. The solid lines correspond to the full catalog, while the dashed lines correspond to the catalog once objects fainter than magnitude $m_H = 23$ are removed. Upper and lower values of the dispersion computed using percentiles with respect to the de-biased distributions are shown according to the legend. The left panel provides the distributions for the field plus cluster member galaxies and the right panel focuses on cluster member galaxies, i.e., those within haloes more massive than $10^{14} M_\odot$.

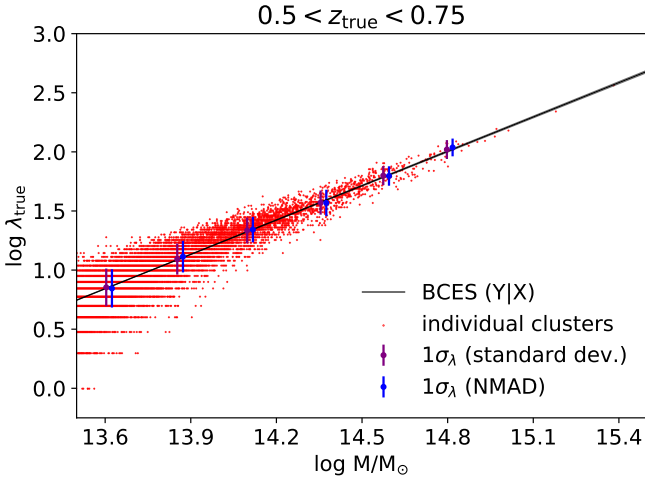


Figure 3: Example of the mass-richness scaling, for the redshift range $z_{\text{true}} = [0.5, 0.75]$. The red dots show the cluster population. The blue points with error bars represent the median richness and scatter computed as the normalized median absolute deviation, while the purple points correspond to the mean richness and the scatter computed as the standard deviation within each bin.

number density distribution of the clusters, $\Sigma(R)$. We therefore investigate the projected radial profiles of the mock clusters by stacking galaxies belonging to clusters in mass and redshift bins. Prior to the stacking, we normalize the projected clustercentric distances by the characteristic radius R_{200} .

In order to study in a quantitative way the mass and redshift evolution of the profiles and compare it to observations from the literature, we use the following approach. We model the profiles by a Navarro, Frenk and White (NFW, Navarro et al. 1996) distribution, as expected from observations (e.g., Carlberg et al. 1997; Lin et al. 2004). However, as we observe a deficit of galaxies in the outskirts of the profiles, we also include a truncation radius, r_{max} , above which the number density of galaxies is set

to zero. The 3D profile of the cluster galaxy space density, n , can thus be written as

$$n(r/R_{200}) = \frac{n_0}{(c r/R_{200})(c r/R_{200} + 1)^2} \mathcal{H}(r_{\text{max}} - r), \quad (2)$$

where $\mathcal{H}$ is the Heaviside step function, n_0 the normalization, $c = R_{200}/r_c$ the concentration, with r_c a characteristic radius, and r_{max} a truncation radius. We fit the stacked normalized number surface density profiles, $\Sigma(R)$, as described by equation (2), using the analytical projection given in Mamon et al. (2010). The parameter space (normalization n_0 , number concentration c , and truncation radius r_{max}) are sampled using a Markov Chain Monte Carlo method, using the algorithm described in Adam et al. (2015).

The left panel of Figure 4 provides the stacked projected profiles of clusters in four redshift and mass bins together with the best-fit models. Overall, the clusters are relatively well described by a truncated NFW model. However, some excess is seen above the best-fit truncation radius, probably due to the fact that each cluster may present a slightly different r_{max} value, while we are introducing blurring in the profile when stacking and only fitting for a unique r_{max}/R_{200} . In addition, the mock clusters present a significantly shallower slope in the center. The best fits are thus slightly biased high in the center, and biased low in the intermediate regions, as seen in the residual. The right panel of Figure 4 gives the marginalized posterior likelihood for the parameter r_{max} versus c . The truncation radius decreases with redshift, being $r_{\text{max}}/R_{200} \sim 1.1 - 1.8$. Such a trend could be due to the fact that r_{max} measures more closely the virial radius, which is defined at higher densities at higher redshifts, leading to radii that will be smaller. However the size of the effect we find is larger than expected. This truncation is not expected from observations, which indicate that the intrinsic cluster number density profile (not counting galaxies in other groups for clusters) extends to over ten virial radii (Trevisan et al. 2017).

The number concentration parameter, is $c \sim 10$ at high masses ($> 10^{14.5} M_\odot$) and increases up to 20 at lower masses (about $10^{13.9} M_\odot$). We note that the values of number concentrations found in these simulations are higher with respect to those

estimated fitting radial number density and stellar mass density profiles of satellite galaxies in observed massive clusters by a factor of about two or more, depending on mass and redshifts (Carlberg et al. 1997; Lin et al. 2004; Collister & Lahav 2005; Muzzin et al. 2007; van der Burg et al. 2014, 2015; Cava et al. 2017). Some of these observational values of concentration and truncation radius (normalized to R_{200}) are reported in the right panel of Figure 4, showing a significant offset with respect to the values estimated from the mock. This discrepancy is similar to that found by Budzynski et al. (2012) comparing number density profiles estimated from SDSS DR7 groups and clusters and predicted profiles from semianalytical modeling of galaxy formation. Indeed, the treatment of the galaxy mergers in the model is shown to impact on the profile shape. When a galaxy becomes a satellite, an analytic estimate of the merger time is made and the galaxy merges regardless of whether or not its host sub-halo can still be resolved. According to the way this merger dynamical timescale is calculated may lead to steeper inner satellite number density profiles in the case of semianalytical models as compared to observed ones. Our main concern here is if this difference could hamper our performance estimation of cluster finders.

As highly concentrated clusters are expected to be more easily identified by cluster finders, this high concentration potentially affects the detections. This may boost high the absolute estimate of the performance, in particular for low S/N objects. However, all the cluster finders are density-based, so their relative performance should not be affected by the higher concentrations. In addition, the truncation of the simulated clusters at 1 to 2 R_{200} facilitates the distinction of the cluster with the background galaxy density, helping the cluster finders limit the dimensions of the clusters on the sky. However, this effect is likely to have a minor impact on the results because the truncation happens at large radii and only marginal effects are visible in the inner part of the clusters once projected along the line of sight.

3.2.4. Cluster galaxy luminosity function

Another important cluster property which can *a priori* affect its detection is its luminosity function. If the galaxy luminosity function in the mock clusters was significantly different from that of the real data, this may impact on the estimate of the absolute performances of the cluster finders. We note however that this is not of major importance for this analysis, in which we are mostly interested to the relative properties of the cluster finders.

We follow the same approach as for the profiles in order to investigate the LF of galaxy clusters within the simulations (see also Ascaso et al. 2015, where the mock galaxy luminosity and mass functions were shown to be in good agreement with observations). We count the number of cluster galaxies in bins of magnitude (in the H -band prior to introducing any noise on the galaxy fluxes), within a projected radius of R_{200} and per Mpc^2 . This is done after selecting clusters within bins of mass and redshift. The LF is then fitted by a Schechter function (Schechter 1976), given by (see e.g., Driver et al. 1994)

$$\Phi(m) = 0.4 \log(10) \phi^* 10^{0.4(m^* - m)(\alpha + 1)} \exp(-10^{0.4(m^* - m)}). \quad (3)$$

As in the case of the galaxy density profile, we fit for the parameters ϕ^* , m^* , and α , which set the normalization, the characteristic magnitude, and the faint-end slope of the population, respectively. Several observational estimates of the cluster LF have shown that a single Schechter function may not reproduce well both the bright and the faint part of the LF, for various reasons (e.g., Popesso et al. 2005; Barkhouse et al. 2007; Yang et al.

2008; Trevisan & Mamon 2017). However it can be used successfully to model its bright part. Here, the Schechter function is not able to describe the mock LF in the faint part (typically $m > m^* + 3$), where a more sophisticated modeling would be necessary. Therefore, we first focus on the bright end of the LF studying the evolution of the parameter m^* . To do so, we perform the fit of equation (3) in the magnitude range limited to $m_{\text{brightest}} + 3$, where $m_{\text{brightest}}$ is the magnitude of the brightest galaxy in the bin we consider. This ensures good modeling of the mock LF in this regime. We check that our best-fit is not sensitive to this magnitude limit. The faint end properties of the LF are addressed as a function of redshift without relying on a model.

The left panel of Figure 5 provides the cluster galaxy LF in two bins of mass (above $10^{14} M_{\odot}$) and five redshift bins (among the twenty considered, from $z = 0$ to 2). We observe that the mock LF are well described by the Schechter function in the bright regime, but that the faint part may require more sophisticated modeling. The right panel of Figure 5 compares the evolution of the best-fit m^* parameter to a passive evolution model derived from Fioc & Rocca-Volmerange (1997), as well as from data taken from the literature. The blue points indicate m_H^* values from de Propriis et al. (1999); Nakata et al. (2001); Ellis & Jones (2004); Lin et al. (2004); Toft et al. (2004); Andreon et al. (2005); Strazzullo et al. (2006); Muzzin et al. (2007); Strazzullo et al. (2010); De Propriis (2017). They were obtained from studies of K-band cluster luminosity functions at different redshifts. We converted the m_K^* values to the H-band using the early-type k-corrections of Mannucci et al. (2001) and the mean rest-frame color for cluster galaxies, $m_H - m_K = 0.26$, obtained as an average of the values provided by Boselli et al. (1997); de Propriis & Pritchet (1998); Ramella et al. (2004), and adopting when needed the transformation to the AB-system $m_{H,AB} = m_H + 1.37$ (Ciliegi et al. 2005). The evolution of the mock is relatively well described by the model and matches well the literature data at redshift larger than 0.3, for the two mass bins considered, but the value of m^* is overall lower by about 0.5 magnitude for the passive evolution model. At lower redshifts, the evolution is stronger with redshift and the mock m^* values are lower than the model and the literature values.

We have also investigated if the performance of the cluster finders could be affected differently according to the way the luminosity function is used in the detection process. While sFoF and HCFA algorithms do not make use of the luminosity function, AMASCFI, AMICO, PZWav and WaZP do. In the case of AMICO, the procedure adopted is fully general and treats the mock as real data. The procedure starts from an initial simple model (built in a blind way) with a luminosity function extracted from all galaxies in the catalog. AMICO is run to define a first set of detections that have been used to refined the cluster model, now introducing a different LF for clusters and field. Finally, AMICO is run with this refined model to derive the final catalog. In the case of PZWav and WaZP, a value of m^* derived from passive evolution model is used to define a constant stellar mass threshold with redshift for detection. However, the dependance of the performance on the m^* cut was tested and found to be negligible. AMASCFI, PZWav and WAZP also use m^* parametrization for richness estimation, but here again richnesses are only used as relative quantities. Therefore, the impact of different uses of the luminosity function by the cluster finders is expected to be negligible on their relative performance.

In addition to the LF itself, we have checked the luminosity differences between the BCG and the central galaxies (i.e., the one coincident with the dark matter halo center in the mocks).

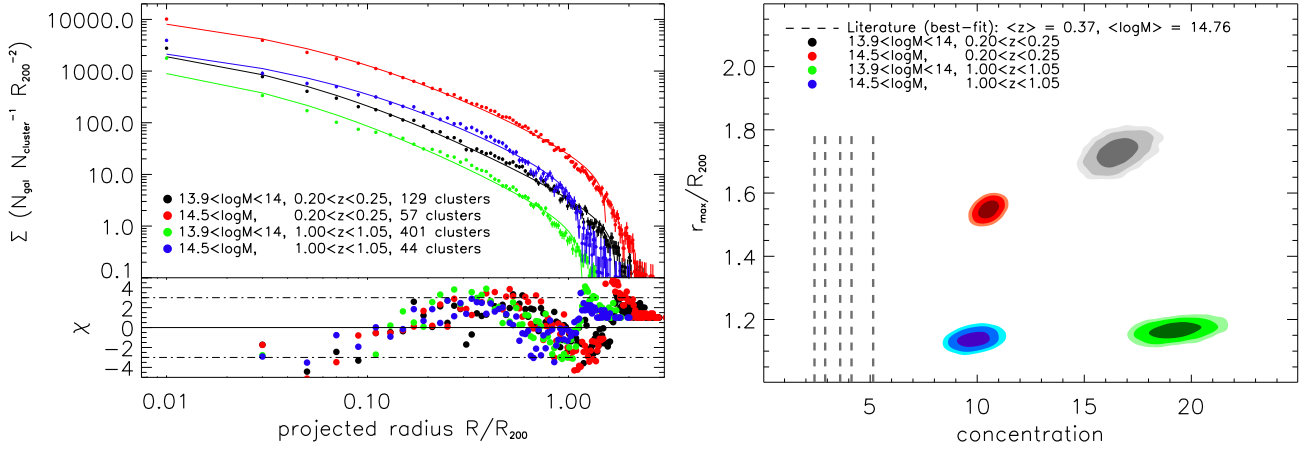


Figure 4: **Left:** stacked surface density (projected) profile of cluster galaxies. The different colors indicate different mass and redshift bins, as indicated in the legend. The solid lines provide the best-fit models of equation (2) in each case. The residual normalized by the error, χ , is also provided. **Right:** posterior likelihood on the model truncation radius parameters $r_{\max}$ and concentration c for each bin, providing the 68%, 95% and 99% C.L. The vertical dashed lines represent the best-fit number concentration cluster observational data from the literature, namely: $c = 2.90 \pm 0.22$ (at a median redshift $\text{med}(z) = 0.04$, and median mass $\text{med}(M_{200}) = 4 \times 10^{14} M_{\odot}$, Lin et al. 2004), $c = 4.13 \pm 0.57$ (at $\text{med}(z) = 0.31$, $\text{med}(M_{200}) = 3 \times 10^{14} M_{\odot}$, Muzzin et al. 2007), $c = 5.14^{+0.54}_{-0.63}$ (at $\text{med}(z) = 1.00$, $\text{med}(M_{200}) = 2 \times 10^{14} M_{\odot}$, van der Burg et al. 2014), $1/c = 0.278 \pm 0.065$ (at $\text{med}(z) = 0.06$, $\text{med}(M_{200}) = 6 \times 10^{14} M_{\odot}$, van der Burg et al. 2014), $c = 2.40 \pm 0.30$ (at $\text{med}(z) = 0.44$, $\text{med}(M_{200}) = 14 \times 10^{14} M_{\odot}$, Annunziatella et al. 2014).

The BCG is coincident with the central galaxy in about 70% of the clusters. This number increases with mass, reaching nearly 100% for the most massive clusters. When the BCG is not the central galaxy, the distance from the BCG to the cluster center (either defined as the central galaxy or the barycenter), is about $0.45 R_{200}$, decreasing by a few percent as mass increases. However, the distribution extends up to around $2R_{200}$ in the low mass clusters. Even when it is not the BCG, the central galaxy is among the brightest members and the magnitude difference with the BCG does not exceed $\Delta m_H \sim 2$, or $\Delta m_H \sim 0.5$ at high mass. The differences between the BCG and the central galaxy can affect the cluster finders to some extent, but we note that no finder relies on the BCG directly. As discussed in Section 4, the associations between the detection based on the BCG and the clusters in the mock could even be missed in a small fraction of the cases, but we have verified that this does not significantly impact the results. We have also checked that the distribution of halo BCG magnitudes in the mocks was in good agreement with observations.

Another important property of the galaxy distribution to be fiducially reproduced by the mocks is the color distribution. We do not focus on that point in this paper since none of the cluster finders participating in the last CFC was relying on galaxy colors. We refer to the work by Ascaso et al. (2015) who found a good agreement in the red sequence properties and the blue valley location between mocks and observation in the redshift range $[0.3, 1.65]$.

4. Mock cluster to detected cluster associations

The assessment of the performance of an algorithm requires associating the candidate clusters and the mock clusters, which are known from the simulation (see e.g., Knobel et al. 2009, and in particular their Figure 3). In this section, we present the methodology developed to perform this association as well as an estimation of the corresponding systematic effects.

4.1. Matching procedures

The association between candidate clusters and mock clusters, or any pairs between cluster catalogs, is a non-trivial task. In order to validate our methodology and test for systematic effects, we have developed three different matching methods. They are hereafter referred to as **geometrical**, **ranking**, and **membership** matching. The matching can generally be performed in two ways, starting from the mock clusters and searching for associated detections, or starting from the candidate clusters and searching for counterparts in the mock. We define the one-way associations as the clusters for which the association has been made in one direction, but not the other one. Similarly, we define the two-way associations as the ones for which the associations are bijective.

4.1.1. Geometrical matching

The geometrical matching method is implemented via the following steps.

1. For each mock cluster, we search for detection counterparts within a volume around the mock cluster. The volume depth along the redshift axis is controlled by the parameter $\Delta z_{\text{match}} = k \sigma_0(1+z)$ where $\sigma_0 = 0.05$ (see Section 3.2.1). We use $k = 4$, i.e., a width of four times the typical photometric error at the given true redshift to ensure avoiding missing matches and minimize false associations. We do not consider any photometric redshift bias or dependence of scatter with redshift, as given in Figure 2. As a result of the photometric redshift uncertainty and inaccuracy, cluster finders for which redshifts are inaccurately calibrated might loose detections that will be considered as impurities, lowering the completeness of the sample. The footprint of the volume, in terms of sky coordinates, is first restricted to the extent of the galaxies belonging to the mock cluster: $\text{RA}_{\min}$, $\text{RA}_{\max}$, $\text{Dec}_{\min}$ and $\text{Dec}_{\max}$. This ensures that the galaxies that are driving the detection are true cluster members and not nearby line-of-sight projected structures. In addition, the volume footprint is re-

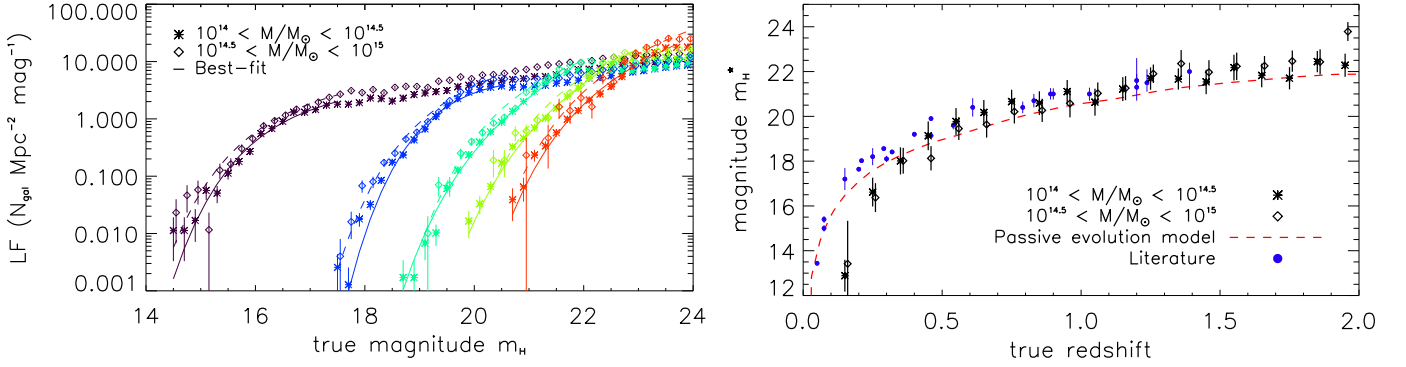


Figure 5: **Left:** stacked LF of cluster galaxies. The different colors indicate different redshift bins, of width 0.1, used to compute them. Only redshift bins centered on 0.15 (purple), 0.55 (blue), 0.95 (cyan), 1.35 (green) and 1.75 (red) are shown for clarity. As indicated in the legend, the star and diamond symbols correspond to the two mass bins, in the range $10^{14} - 10^{14.5} M_{\odot}$ and $10^{14.5} - 10^{15} M_{\odot}$, respectively. The solid and dashed lines provide the best-fit models of equation (3) in the bright magnitude regime, in the low and high mass bins, respectively. We note that in the high mass bin, the number of clusters per bin may be less than 10 at redshifts larger than 1.5, and reaches 2 in the last bin. **Right:** redshift evolution of the parameter m_H^* , for each mass bin using similar symbols, and comparison to the passive evolution model, as the red dashed line, from [Fiocc & Rocca-Volmerange \(1997\)](#), and calibrated using the work by [Lin et al. \(2006\)](#). The blue points indicate m_H^* values from the literature (see text). The error bars provide the standard deviation of the posterior distribution of the parameters m_H^* , but we stress that the distributions are generally non gaussian and non symmetric (see [Ricci et al. 2018](#), for a detailed discussion on this topic).

stricted to be within θ_{200} of the mock cluster, the angular radius corresponding to R_{200} , given the mock cosmological parameters. In the case of massive and nearby clusters, this last condition is more restrictive than the first one. However, as redshift increases and mass decreases, the number of cluster galaxies remaining above the mock flux limit drops, and for a given cluster, all the mock cluster galaxies are eventually enclosed within θ_{200} . In this case, this secondary constraint becomes ineffective with respect to the first one.

2. In the case of multiple counterparts within the volume, we define the matched cluster as the one which is the closest (projected on the sky) to the mock cluster. Nevertheless, we record the total number of possible matches for all mock clusters, as they correspond to fragmented detections.
3. We repeat the first step (search for counterparts in the volume around the cluster), using the candidate clusters as the reference and searching for mock counterparts. While the redshift criterium is symmetric and remains the same, it is not the case for the projected area because the detection algorithms do not provide a characteristic radius of the detected objects. Therefore, a mock cluster is associated with the detection if it is at a projected distance that is lower than its own θ_{200} .
4. We repeat step 2 with candidate clusters as the reference.
5. By comparing mock and cluster detection counterparts, we identify mock clusters and detected clusters for which the association is identical both ways.

This method allows us to define both the one-way and the two-way associations. In case a mock cluster is associated with a multiple number of detections, this indicates that fragmentation has occurred and this is an important quality assessment of a cluster finder. Similarly, detected clusters that are matched to multiple mock clusters correspond to over-merging events. The two-way geometrical matching is taken as the baseline method in the present paper.

4.1.2. Ranking matching

The ranking matching method follows the same initial condition as the geometrical matching (first step: search for counterparts in the volume around the cluster). However, instead of

performing the matching both ways, it associates candidate clusters to mock clusters after ranking them by decreasing mass, as provided from the mock catalogs, and richness, as provided by the cluster finders, respectively. The richest detected clusters are then matched to the most massive mock clusters, and subsequently removed from the list. If two or more clusters have the same richness within the association volume, the nearest one to the mock cluster center is selected. Because detected and mock clusters are subtracted from the cluster list as they are matched to one another, this matching procedure is bijective by construction and thus corresponds to a two-way matching. It cannot, therefore, be used to address fragmentation and over-merging issues. The ranking matching follows the idea that the most massive clusters, i.e., the richest ones, are the first ones detected.

4.1.3. Membership matching

The geometrical matching and the ranking matching do not directly rely on the cluster member galaxies. In contrast, the third method we developed, hereafter membership matching, consists in defining the associations using the galaxies that are detected as cluster members by the algorithms. Because not all the detection algorithms provide the galaxy membership information, this method is only used as a crosscheck (see Table 1 for the algorithms that provide the membership). The main steps of the procedure are summarized as follows.

1. For each mock cluster, we search for detection counterparts within the volume as defined in the case of the geometrical matching.
2. If matches are found, we define the fraction of common galaxies between the candidate cluster and the mock cluster, with respect to the mock cluster (i.e., the success rate) as

$$f_{\text{com, mock}} = \frac{\sum_i P_i^{\text{match}}}{N_{\text{gal, mock}}}, \quad (4)$$

where P_i^{match} are the cluster membership probabilities of the galaxies which are indeed true cluster members and $N_{\text{gal, mock}}$ is the number of galaxies that belong to the cluster according to the mock catalog. The matches are then accepted if

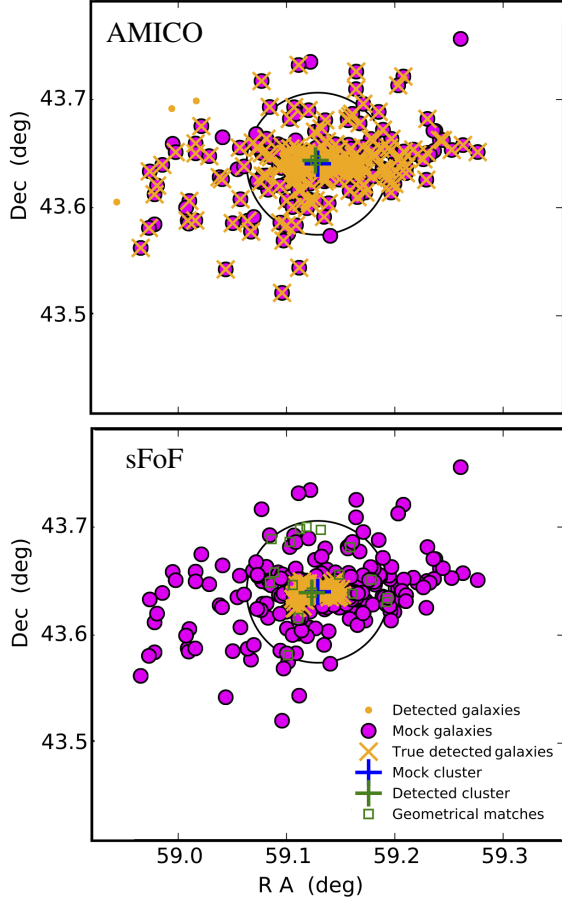


Figure 6: Illustration of the matching procedure performance in the case of a test cluster ($z = 0.285$, $M = 10^{14.04} M_{\odot}$), for two algorithms for which the galaxy membership probability was available (AMICO on the top and sFoF on the bottom panel). The true cluster galaxies, known from the mock, are given in purple. The cluster galaxies identified as members by the cluster finders are given in yellow (only galaxies with $P_i > 0.25$ are shown for clarity), with a cross on top in case they are true cluster members. The target mock cluster is given as a blue cross. All the candidate clusters that lie within the association volume are given as green squares, with a cross on top of the one that is the best match. The black circle represents θ_{200} . In the case of sFoF, 17 one-way geometrical associations would be possible (number of green squares), highlighting a large fragmentation rate.

$f_{\text{com, mock}} > f_{\text{cut}}$, with f_{cut} a free parameter of the matching. In the case of possible multiple matches, the candidate cluster with the largest $f_{\text{com, mock}}$ is defined as the best match, but we still record the total number of possible matches for fragmentation estimates. Since the definition of membership probability is different for the different detection algorithms (e.g., a large number of galaxies dominated by low probability objects due to large photometric redshift error, or few galaxies with binary membership probabilities), the minimum fraction of common galaxies was taken to be $f_{\text{cut}} = 0$, i.e., any cluster with non-zero $f_{\text{com, mock}}$ was considered as a possible match.

3. We repeat steps 1 and 2, using candidate clusters as the reference and searching for mock counterparts. This time, the fraction of common galaxies is defined with respect to the candidate clusters as

$$f_{\text{com, det}} = \frac{\sum_i P_i^{\text{match}}}{\sum_i P_i^{\text{all}}} \quad (5)$$

The minimum fraction of common galaxies was also taken to be $f_{\text{cut}} = 0$ in this direction.

4. By comparing mock and cluster detection counterparts, we identify clusters for which the association is bijective.

Because of the low value of f_{cut} , the one-way associations are redundant with the one-way geometrical matches. On the other hand the two-way associations rely on bijective galaxy associations instead of distance.

In Figure 6, we illustrate the different matching procedures using one test cluster at $z = 0.285$ and $M = 10^{14.04} M_{\odot}$, in the case of two cluster finders that provide the membership probabilities (AMICO and sFoF). While all the matching methods agree on the detection of this mock cluster for both algorithms, the geometrical and membership matching methods show that fragmentation is important for sFoF in this test case.

4.2. Purity, completeness, fragmentation and over-merging

The performance of an algorithm is related to the quality of the cluster catalog that it produces. It is generally quantified in terms of completeness (i.e., the number of detected clusters normalized by the number of clusters in the simulation) and purity (i.e., the number of true detections normalized by the overall detection number), which can be expressed as a function of cluster properties such as redshift, mass or richness. With the matching results in-hand, it is straightforward to define the one-way purity and completeness as

$$\begin{cases} P_1(\Delta\vec{\mu}) = \frac{\mathcal{A}_1[N^{\text{det}}(\Delta\vec{\mu}) \rightarrow N^{\text{true}}]}{N^{\text{det}}(\Delta\vec{\mu})} \\ C_1(\Delta\vec{\mu}) = \frac{\mathcal{A}_1[N^{\text{true}}(\Delta\vec{\mu}) \rightarrow N^{\text{det}}]}{N^{\text{true}}(\Delta\vec{\mu})} \end{cases} \quad (6)$$

where N^{true} is the number of mock clusters and N^{det} is the number of detected clusters. The quantity $\Delta\vec{\mu}$ stands for a bin in terms of cluster parameters such as $\vec{\mu} \equiv (M, z, \lambda)$, so that $\mathcal{A}_1[N^{\text{det}}(\Delta\vec{\mu}) \rightarrow N^{\text{true}}]$ is the number of mock clusters in the bin $\Delta\vec{\mu}$ that are associated with a detection. We note that while the completeness can be estimated as a function of true mass, redshift or richness, purity cannot be expressed in a given mass bin unless the measured mass-richness relation is used. Similarly, the two-way purity and completeness (available for the geometrical and membership matching procedures), are defined as

$$\begin{cases} P_2(\Delta\vec{\mu}) = \frac{\mathcal{A}_2[N^{\text{det}}(\Delta\vec{\mu}) \longleftrightarrow N^{\text{true}}]}{N^{\text{det}}(\Delta\vec{\mu})} \\ C_2(\Delta\vec{\mu}) = \frac{\mathcal{A}_2[N^{\text{true}}(\Delta\vec{\mu}) \longleftrightarrow N^{\text{det}}]}{N^{\text{true}}(\Delta\vec{\mu})} \end{cases} \quad (7)$$

in which case we impose that the associations be bijective.

We note that impurities do not only correspond to spurious objects, such as improperly identified structures along the line-of-sight, or unmatched fragments of larger clusters. They may also correspond to clusters for which the mass is below the threshold of the mock cluster catalog constructed as defined in Section 3.1.2. This point is discussed further in sections 5 and 6.

We define the N-fragmentation rate as the fraction of mock clusters for which more than N-associations to detected clusters are possible. Similarly to completeness and purity, it can be

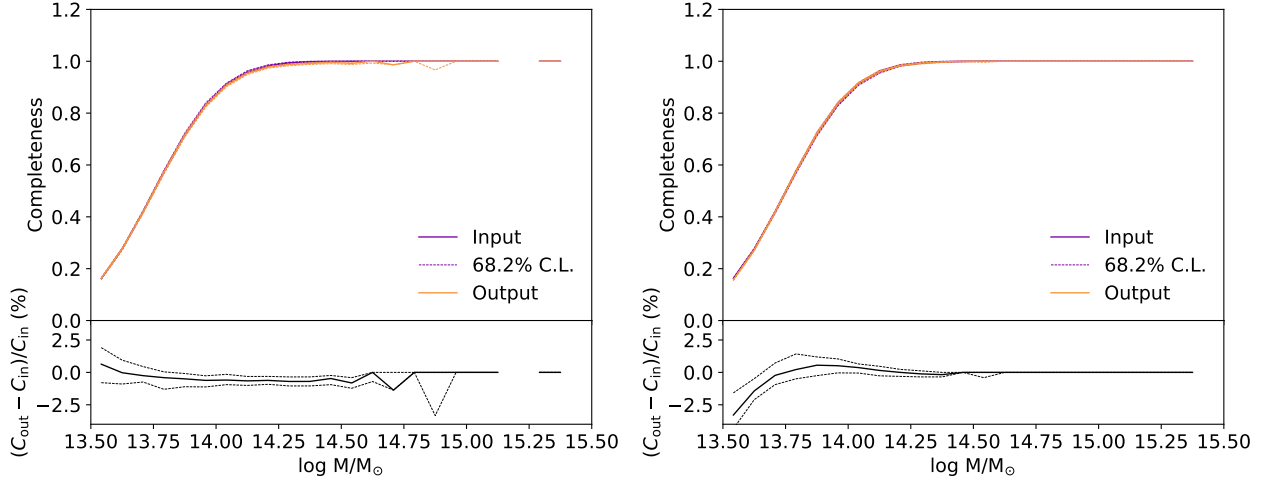


Figure 7: Comparison between the input completeness, as a function of mass, and the recovered completeness for the **geometrical** matching (left) and the **ranking** matching (right). The normalized residual is provided as the bottom plot. The dashed line provide the statistical uncertainty on the bias, computed as the 68.2% statistical limit over all the Monte Carlo realizations.

expressed as a function of mass, redshift and richness. The N-over-merging rate is given by the fraction of detected clusters for which more than N-associations to mock clusters are possible. Having $C_1 \simeq C_2$ and $P_1 \simeq P_2$ is also an indication of low fragmentation and low over-merging.

For each detected cluster that is matched to a mock cluster, we compute the redshift difference between the true cluster redshift and the recovered one. We additionally compute the projected sky coordinate offsets (see Section 5.4 for further discussions and the use of these quantities).

4.3. Systematic effects

The matching procedure is ambiguous in the sense that the detection of a cluster has to be addressed based on somewhat arbitrary criteria. In addition, galaxy clusters are extended objects, with internal structure that varies among the cluster population. As a consequence, the matching itself is not immune to artifacts that are reflected in the selection function.

In order to estimate the systematic effects associated with the matching, we use the following procedure. We construct a new mock cluster catalog by removing clusters randomly, following a realistic input completeness that we define, C_{in} , based on the results described in Section 5 (typically using the overall results of the different algorithms). As shown in Figure 7, the completeness is described by an error function (erf) with characteristic mass of $10^{13.75} M_{\odot}$ and with 0.21 dex width. We also introduce noise in the sky and redshift coordinates of the clusters, modeled by a Gaussian distribution, representative of the results of Section 5.4, as a function of mass and redshift. The redshift standard deviation is set to $0.03(1 + z_{true})$ at $M = 10^{13.5} M_{\odot}$, and evolves as $1/\sqrt{M}$. The position standard deviation is fixed to 0.25 arcmin. Since the **ranking** matching also relies on the richness estimates, we introduce an extra scatter of 0.2 dex in the richness, based on the statistical scatter as measured in Section 5.3. The new constructed cluster catalog mimics the properties of the catalog produced by the detection algorithms and it is matched to the original catalog to estimate the biases induced by the matching procedure on the completeness. As each newly generated mock catalog is only a noisy representation of

the mean catalog, this procedure is repeated by generating 100 Monte Carlo realizations of such a catalog. The mean recovered completeness, C_{out} , allows us to measure biases and the dispersion from the Monte Carlo realization used to estimate statistical errors. This method is applied only to the **ranking** matching and the **geometrical** two-way matching, because it cannot handle the galaxy membership required by the **membership** matching procedure. It is therefore only used as a cross check.

Figure 7 provides the comparison between the input completeness and the recovered one, for the **geometrical** two-way and the **ranking** matching, as a function of mass. For the **ranking** matching, we observe a mass dependent bias, reaching up to 3% at $10^{13.5} M_{\odot}$. This is due to the fact that low mass clusters can be missed when there is a more rich, competing cluster, in their surrounding. The overall systematic effects due to the matching procedure remain below 3%.

5. Results and performance

The blind detection of galaxy clusters using the algorithms presented in Section 2 on the mock catalog (discussed in Section 3) was followed by the association between detections and mock clusters (Section 4). In this section we address the detection performance expected for *Euclid*. The overall performance is summarized in Table 2.

5.1. Comparison criteria

The competing cluster finders do not provide cluster catalogs down to the same detection limit. Additionally, the S/N of the detections are only available for a subset of the cluster finders. Nevertheless, a ranking of the detections, according to their reliability, is provided for all of them. In most cases, the ranking is performed according to S/N but for cluster finders not providing S/N ranking is evaluated according to richness. In order to compare the global performance of the algorithms, we impose a minimum purity on the cluster detection catalogs. This is done by removing the detection with lowest ranking, until the minimal purity is reached for the overall sample. This is illustrated in Figure 8, where we compute the purity as a function of completeness by limiting the detection catalog to the most reliable objects

Table 2: Summary of the performance of the cluster finders, as applied on the mock. Note that in the case of WaZP, the median centroid offset is zero because most detections have been assigned to the true mock cluster central galaxies.

Finder	Mean C for a $P = 0.8$ (at $M > 10^{14} M_{\odot}$, $z_{\text{true}} < 2$)	Mean P for a $C = 0.8$ (at $M > 10^{14} M_{\odot}$, $z_{\text{true}} < 2$)	Mean statistical M - λ_{det} scatter (at $M > 10^{14} M_{\odot}$, $z_{\text{true}} < 2$, in dex)	Redshift RMS / $(1 + z_{\text{true}})$ (at $M > 10^{14} M_{\odot}$)	Centroid offset (mean, median, RMS) (at $M > 10^{14} M_{\odot}$, arcmin)
–					
AMASCFI	65%	Not reached	0.27	0.025	(0.44, 0.36, 0.33)
AMICO	91%	93%	0.19	0.015	(0.22, 0.19, 0.19)
HCFA	55%	Not reached	0.15	0.021	(0.46, 0.33, 0.45)
PZWav	83%	86%	0.16	0.020	(0.27, 0.21, 0.45)
sFoF	57%	66%	0.17	0.017	(0.16, 0.11, 0.15)
WaZP	83%	83%	0.18	0.016	(0.10, 0.00, 0.26)

up to a given rank, which is varied from unity (i.e., only the best detection) to the total number of detections in the catalogs (i.e., until the least reliable detections are included). To illustrate the expected *Euclid* data usage in a cosmological context, the target mock catalog is restricted to redshifts $z_{\text{true}} < 2$ and masses $M > 10^{14} M_{\odot}$. The completeness is thus computed using only the objects satisfying these limits, while the purity reflects the content of the full detection catalogs. The completeness will be discussed in more detail in Section 5.2.

Since the full mock cluster catalog only includes objects with masses larger than $10^{13.25} M_{\odot}$, detections at masses below this limit will appear as impurities even though they can correspond to real groups. To estimate the importance of this effect, considering the mass range we adopted, we produce a perfect detection-like catalog using the true mock catalog. The clusters are ranked by richness, accounting for redshift evolution as detailed in Section 3.2.2 and including an extra scatter of 0.2 dex that mixes the order of the detections as expected for real cluster finders. We then consider the detected clusters to be true detections (i.e., they have been matched to mock clusters) only if their mass is above a given threshold, M_{cut} . This threshold thus mimics the mass limit that we define when constructing the mock catalog ($10^{13.25} M_{\odot}$ for our baseline). We reproduce Figure 8, in the same range of mass and redshift for completeness computation, but for different values of the threshold M_{cut} . The results are shown in Figure 9, where we can see that while the mock catalog mass threshold remains close enough to the detection limit, the loss of purity is small. A bias of $\lesssim 5\%$ is observed up to mass cut as large as $M_{\text{cut}} = 10^{13.6} M_{\odot}$ (compared to the limit of $10^{13.25} M_{\odot}$ we have used). Based on the detection limit of the cluster finders (see Section 5.2), the bias is expected to be about 1 – 2% at most. In addition to this crosscheck, the calculations of Figure 8 were repeated for a catalog mass cut of $10^{13} M_{\odot}$ (instead of $10^{13.25} M_{\odot}$), showing no significant differences with the baseline choice.

Prior to further post-detection analysis, it was necessary to restrict the cluster detection catalogs to a common detection significance. This was done in order to compare the performance of the different algorithms, particularly in terms of completeness, given the heterogeneous nature of the detection catalogs provided for the CFC. Cluster cosmology requires a well-defined cluster catalog and we thus restrict the cluster finder samples to ranks good enough to ensure that the mean sample purity is equal to 80% over the mass and redshift range given by $M > 10^{14} M_{\odot}$ and $z_{\text{true}} < 2$. This corresponds to defining a threshold at 80% purity in Figure 8 and excluding clusters with ranks beyond the corresponding limit. This cut at $P = 0.8$ also corresponds to the limit for which the purity starts to drop rapidly with limited improvement for the completeness. The mean completeness of the corresponding catalog is given in Table 2, for this purity threshold, as well as the purity for a completeness threshold of 80%.

This baseline limit does affect the overall performance presented in the following, and we discuss in Section 6 how they change according to this choice. We note that for real observations, such a purity cut cannot easily be applied. However, calibration of the purity as a function of the catalog S/N limit could be done with help from the *Euclid* deep survey or using external data at various wavelengths. The completeness performance will be affected when changing the S/N threshold of the trimmed catalog, as can be seen in Figure 8.

In the following subsections, we address the performance of the six cluster finder algorithms and compare their behaviors in terms of cluster completeness, purity, dispersion introduced by the detection in their mass proxy, the redshift and centroid recovery, fragmentation and over-merging. The selection of *Euclid* algorithms was done by prioritizing high redshift objects ($z \gtrsim 1$) at high mass ($M \gtrsim 10^{14} M_{\odot}$), those that are expected to carry most of the statistical power in constraining cosmological parameters of interest to *Euclid* (e.g., the dark energy equation of state, Sartoris et al. 2016).

5.2. Completeness and purity

The completeness as a function of mass and redshift, is shown in Figure 10 for all six cluster finders. It is computed after removing the least reliable detections to reach a mean purity of 80% for all algorithms. As it is not possible to provide error bars in Figure 10, we also show the completeness as a function of redshift in different mass bins in Figure 11. The error bars are computed using binomial statistics according to the Wilson score interval approach based on the number of detections in each bin. We stress that the mock catalog is the same for all detection algorithms, therefore these error bars only reflect the absolute statistical uncertainty and should not be considered when comparing the differences between the cluster finders.

All six detection algorithms provide high levels of completeness. As expected, higher mass systems are better detected than lower mass ones, regardless of the cluster finder. Clusters at masses $M > 10^{15} M_{\odot}$ are all detected, except for two of them that are missed by HCFA at $z \approx 0.25$ and 0.6. In the case of AMICO and PZWav, up to 50% of the clusters are recovered down to masses of $10^{13.5} M_{\odot}$ at low redshift. We observe a different redshift evolution for the completeness for the various finders given their different sensitivity to the mock properties. Nonetheless, a drop of up to 10% in cluster completeness is seen at $z \sim 0.5$ for most finders (except AMICO), and it could correspond to a feature in the photometric redshifts discussed in Section 3.2.1. The redshift evolution of the completeness could also be affected by issues related to the ranking because some codes may have given high rank preferentially in a given redshift bin and could lose some detections at specific redshifts.

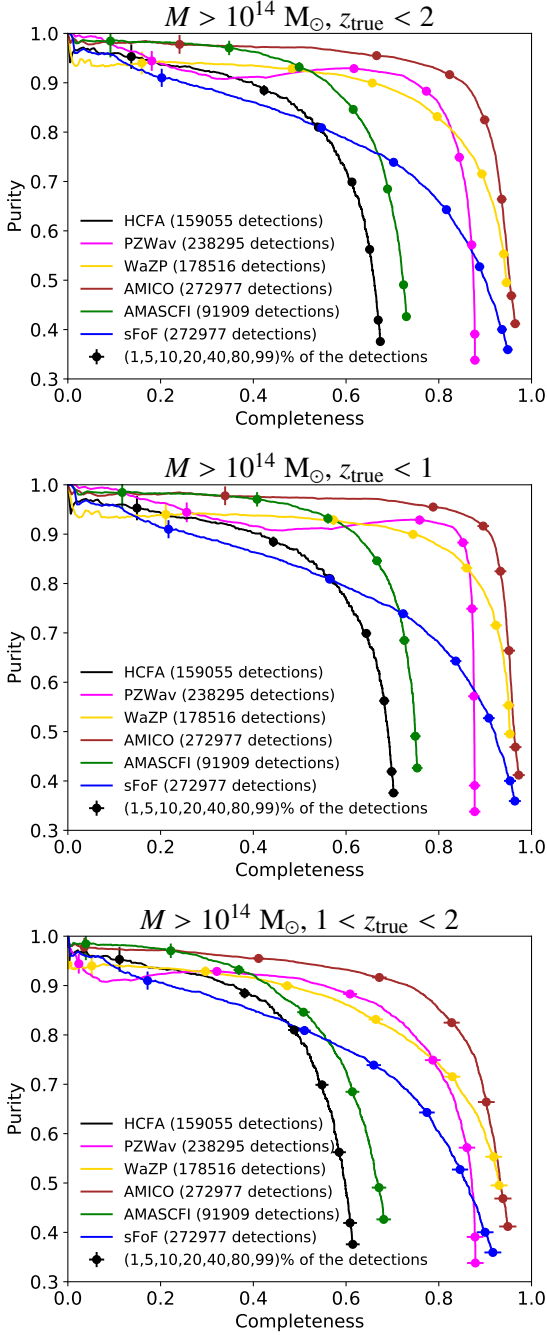


Figure 8: Purity as a function of completeness, given for all cluster finder catalogs as a function of the ranking of the detections. As the number of considered detections above a given rank increases, the detection properties evolve from the high purity low completeness regime to the low purity high completeness regime. The dots provide the percentage of considered detections at a given coordinate on the curve, the total number of detections being given in the legend for each finder. The top panel only accounts for mock clusters in the range $z_{\text{true}} < 2$ and $M > 10^{14} M_{\odot}$ when computing the completeness. For illustration, we also provide the same figure in the range $z_{\text{true}} < 1$ and $1 < z_{\text{true}} < 2$ in the bottom panels.

In order to assess the quality of the purity and compare it among cluster finders, we restrict the original detection catalogs to a fixed mean completeness for $0 < z_{\text{true}} < 2$ and $M > 10^{14} M_{\odot}$ (see also the discussion of Section 5.1, Figure 2 and Table 2). We first choose a mean completeness of 65%, which is reached by

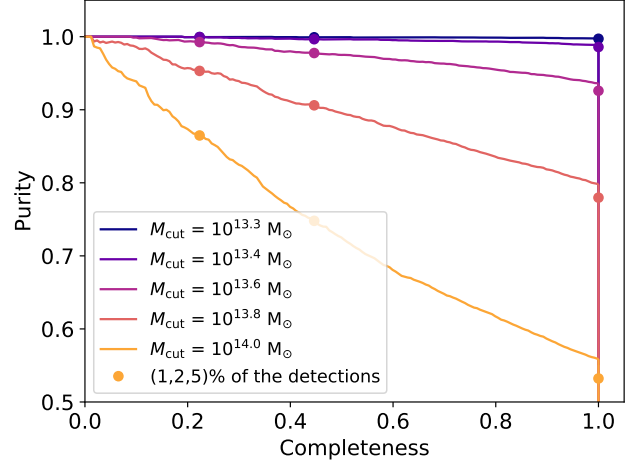


Figure 9: Purity versus completeness constructed for an ideal detection catalog. The different curves correspond to different mass cuts assumed for the mock catalog. The redshift and mass ranges considered are the same as in the top panel of Figure 8.

all the codes, and investigate the evolution of purity as a function of redshift (Figure 12 top panel). We can see that the purity evolves very differently with redshift for the different algorithms. AMICO, PZWav and WaZP provide pure samples up to high redshift, except for $z \lesssim 0.25$, where the purity drops. This might be due to a strong dependence on the photometric redshift quality of these algorithms, as the outlier fraction in the photometric redshifts strongly increases at $z < 0.25$. It could also be caused by the unavoidable fragmentation of very rich clusters, with secondary fragments being counted as impurities. Additionally, in this regime, the mass detection limit of these algorithms (see Figure 10) may be sufficiently low such that some detected clusters are below the mass threshold of the mock and are thus counted as impurities (see also the discussion in Section 5.1). On the other hand, AMASCFI, HCFA and sFoF catalogs are more pure at low redshifts, but their purity smoothly declines as redshift increases. We also analyzed the redshift evolution of the purity for detection catalogs trimmed to reach 80% completeness. Since HCFA and AMASCFI do not reach this completeness limit, they are not considered here. The comparison of the four resulting algorithms is shown in Figure 12 (bottom panel). We can see that the purity is still above about 90% and relatively flat in the redshift range $[0.25, 2]$ for AMICO and PZWav. The performance of WaZP is slightly lower, but still very good at these redshifts (about 80%-90%). The purity of the algorithm sFoF, on the other hand, smoothly declines as redshift increases in this regime. At low redshifts (below 0.25), sFoF remains relatively flat with a purity of about 85%, while the purity of AMICO, PZWav and WaZP slightly decline, down to about 60% to 70%. Nevertheless, we stress that at low redshifts, the mean completeness of these algorithms is much higher than the mean completeness that we impose for this comparison exercise, due to redshift variations. Thus, contamination may arise from objects at the detection limits of the catalog.

5.3. Mass-richness scaling relation

Detections matched to mock clusters were used to investigate the quality of the richness provided by the cluster finders. The

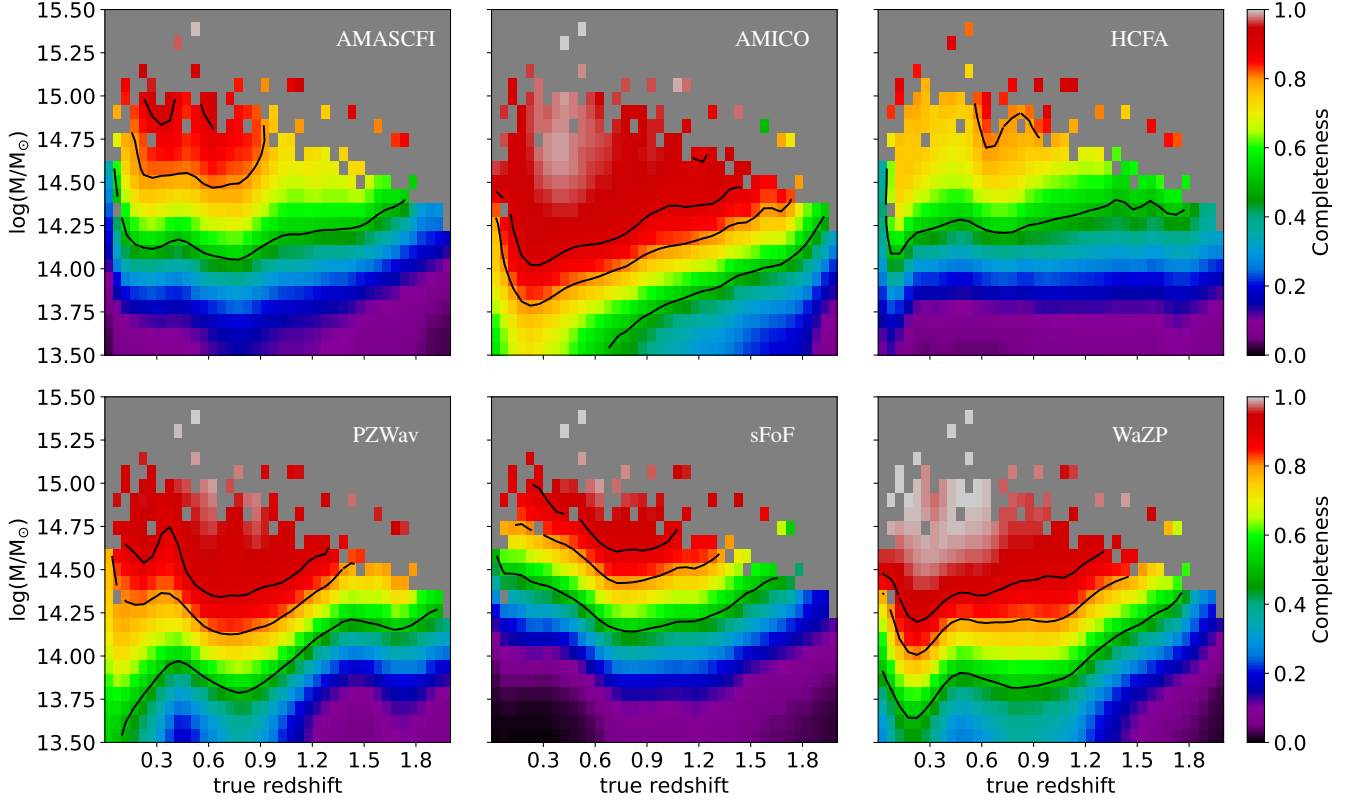


Figure 10: Completeness as a function of mass and redshift for the six final CFC algorithms. The black lines corresponds to completenesses of 50, 80 and 90%. The pixel size corresponds to about $\Delta \log(M/M_\odot) = 0.08$ and $\Delta z = 0.05$, but the image was smoothed with a Gaussian kernel to a resolution of $0.20 \log(M/M_\odot) \times 0.25$ (FWHM) for display purposes. The mask is applied where the number of clusters per pixel is zero before smoothing. Each catalog of detections has been trimmed to the most reliable detections, ensuring a mean purity of 80% in the range $M > 10^{14} M_\odot$ and $z_{\text{true}} < 2$.

aim of this section is to qualitatively compare the goodness of the mass proxy derived from the various algorithms, without going to a detailed characterization of the mass-observable relation, which will be addressed in future work. To do so, we use the full catalogs of true detections (matched clusters), considering all available ranks, because of higher statistics and higher completeness, but the results are not sensitive to this choice because only objects well above the detection limit (see below) are used. The analysis described in Section 3.2.2 is reproduced using a mass proxy as provided by the cluster finders. While the richness definition varies from algorithm to algorithm, they are all expected to scale with mass. In order to have comparable numerical values, the richness values provided by the respective algorithms were normalized to the minimum richness. However, this only affects the normalization of the mass-richness scaling, which is not a concern for the present comparison. To mitigate effects induced by Malmquist bias, resulting from the detection limit of the algorithms, we fit the scaling relation with a power law of slope s : $\lambda_{\text{det}} \propto M^s$ in the range where the median richness is above the detection limit by more than three times NMAD.

Figure 13 illustrates the scaling relation and its best-fit for the six detection algorithms, in the redshift range [0.5, 0.75]. This figure can be directly compared to Figure 3. We can see that all cluster finder mass proxies have different definitions, which scale in different ways with mass. For instance, the mass proxy scales with different slopes: 0.65, 0.61, 1.27, 0.51, 0.96 and 0.78 for AMASCFI, AMICO, HCFA, PZWav, sFoF and WaZP, respectively. The slope remains stable independently of the num-

ber of matched clusters we consider. The mass proxy is discrete in the case of AMASCFI and sFoF, while it is continuous for the other ones. Even an ideal mass proxy derived from the true galaxy members in mock catalogs would have a significant scatter with respect to the true halo mass, as we have shown in Section 3.2.2. As our aim is to measure the performance of the algorithms, we subtract this intrinsic scatter from the measured one. The statistical scatter induced in the detection process is computed as $\sigma_{\text{det}} = \sqrt{\sigma_{\text{meas}}^2/s_{\text{meas}}^2 - \sigma_{\text{int}}^2/s_{\text{int}}^2}$, where σ_{meas} is the overall scatter of richness at a given mass as measured after the detections are performed and s_{meas} the associated slope, and σ_{int} is the intrinsic scatter as discussed in Section 3.2.2 with s_{int} the associated slope. We note that the different mass dependence, given by the slope s , between all the algorithms, is accounted for when computing the scatter. The scatter σ_{det} was measured for different mass bins and redshift bins. We provide its median value for masses above $10^{14} M_\odot$ in Table 2. Once normalized by the slope, HCFA presents the smallest scatter, with a median value of 0.15 dex. Most algorithms perform reasonably well, with $\sigma_{\text{det}} \approx 0.18$ dex, going up to 0.27 dex for AMASCFI (see, e.g., Andreon 2015, for comparison with low scatter richness-based mass proxies).

5.4. Redshift and centroid reconstruction

The matched detections are used to compute the differences between the recovered cluster redshifts and the true redshifts.

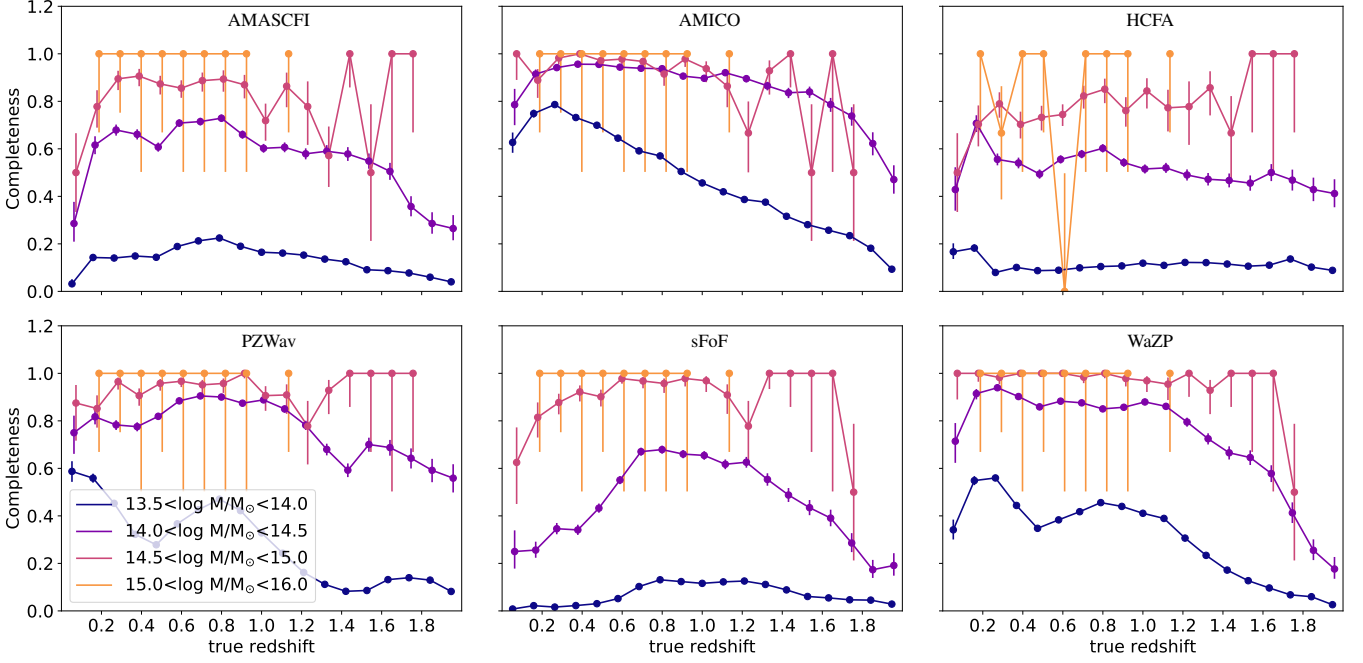


Figure 11: Completeness as a function of redshift in four mass bins, for the six final CFC algorithms. The four colors correspond to different mass bins, respectively $[10^{13.5}, 10^{14}]$, $[10^{14}, 10^{14.5}]$, $[10^{14.5}, 10^{15}]$ and $[10^{15}, 10^{16}] M_{\odot}$. The error bars represent the 68% confidence interval, following the Wilson score interval approach. Each catalog of detections has been trimmed to the most reliable detections insuring a mean purity of 80% in the range $M > 10^{14} M_{\odot}$ and $z_{\text{true}} < 2$.

Figure 14 shows the distribution of this redshift difference for each algorithm and for different mass intervals. We also present the standard deviation of this distribution as a function of mass in two redshift bins ($0 < z_{\text{true}} < 1$ and $1 < z_{\text{true}} < 2$). Because of the matching procedure, this difference is limited to $(1 + z_{\text{true}})\sigma_{z_{\text{phot}}}$ with a typical standard deviation of $\Delta z/(1 + z_{\text{true}}) \sim 0.02$ at $M > 10^{14} M_{\odot}$ (compared to 0.2 for the depth of the cylinder along the line of sight used for the matching procedure). We observe that the redshifts of more massive clusters are slightly better recovered than lower mass ones for all detection algorithms (typically an increase in the scatter by a factor of two from $10^{14.5}$ to $10^{13.5} M_{\odot}$), because more galaxies are accessible to perform the redshift estimates. Only a very small redshift dependence is visible once normalized by $(1 + z_{\text{true}})$. While the redshift difference distribution is always compatible with zero, we can observe a small mass-independent redshift bias for all cluster finders, which reflect the photometric redshift bias (see Section 3.2.1). The standard deviation of the redshift difference is shown in Table 2 for massive clusters ($M > 10^{14} M_{\odot}$). AMICO and WaZP are the ones that perform the best, reaching $\Delta z/(1 + z_{\text{true}}) = 0.015$, but all the algorithms demonstrate a relatively tight performance, with a maximum of 0.025 for AMASCFI.

As for the redshifts, we compute the centroid offset between the detection centers and the mock cluster catalog centers (see Figure 15, where the difference is shown in terms of declination offset). We do not observe significant evolution of the distribution with mass or redshift. This is likely due to the fact that more massive clusters are better detected, but also more extended, which compensates the precision in the centroid recovery. Similarly, nearby clusters are more extended, but also better detected than their high redshift counterparts. The mean, median and standard deviation values of the centroid offset distributions are reported in Table 2 at $M > 10^{14} M_{\odot}$, and corresponds to a typical offset of 0.2 arcmin. Using the mean of the centroid off-

set distribution as a quality indicator, the algorithms WaZP, sFoF and AMICO present the best coordinate determinations, down to 0.10 arcmin. The highest mean of the angular offset distribution is 0.46 arcmin, for HCFA. We note that in the case of WaZP, a large fraction of clusters are detected with zero offset as WaZP defines the center as the peak of a density map, or as the brightest member if found within 50 kpc of the density peak.

5.5. Fragmentation and over-merging

We compute the fraction of fragmented clusters and the fraction of over-merging according to the definitions given in Section 4.2: the number of fragments and overmerged structures are defined as the number of possible one-way associations with respect to the mock clusters and the detected clusters, respectively. We note that our definition of fragmentation and overmerging is tailored to the cluster definition in the mocks, and is thus somewhat arbitrary. We compute the over-merging rate only in the case of two or more overmerged structures. In the case of fragmentation, we compute it for 2 fragments or more, but also for 5 fragments or more (this corresponds to the fraction of mock clusters for which more than one, or more than four, associations are possible).

The N-fragmentation rate and the N-over-merging rate are given in Figures 16 and 17 as a function of mass for different redshift bins. All cluster finders present a fragmentation rate that increases with mass and decreases with redshift. This is expected as more substructures become accessible as the mass increases. Similarly, substructures cannot be resolved at higher redshift as fewer galaxies are accessible. Fragmentation could also arise because a code detects lower mass structures that are inside the matching radius of a larger system. In this sense, a code that is more complete at lower masses will automatically have a higher fragmentation rate. We observe significant differences between the different cluster finders. While AMASCFI,

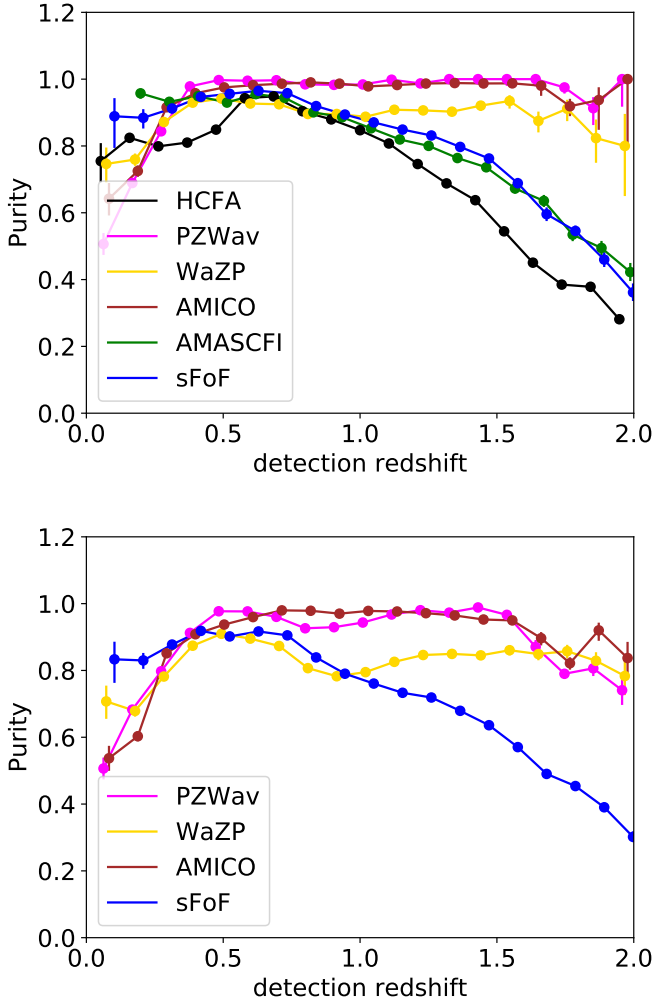


Figure 12: Purity as a function of redshift for a mean completeness of the detection catalogs set to 65% (top panel) and 80% (bottom panel) in the range $0 < z_{\text{true}} < 2$ and for masses $M > 10^{14} M_{\odot}$. AMASCFI and HCFA do not reach the 80% completeness level and are thus not represented in the bottom panel. Error bars are computed as in Figure 11.

HCFA and PZWav present low fragmentation rates, AMICO and WaZP present a higher fragmentation rate, and sFoF tends to fragment clusters up to very low masses and with a large number of fragments (more than 5 fragments are common, as already illustrated in Figure 6 for this cluster finder). The over-merging rate remains at the level of about 10% and is relatively constant in mass. For most algorithms, we observe an increase in the over-merging rate for lower mass systems, which is likely due to the fact that these objects are more abundant. The over-merging rate is comparable for all detection algorithms, even if AMASCFI, HCFA and sFoF algorithms are slightly more affected than AMICO, PZWav and WaZP. We note that similarly to the fragmentation rate, the overmerging rate is affected by the matching procedure. Indeed, two objects that are correctly detected by a cluster finder, but aligned along the line-of-sight within an overlapping matching cylinder, will appear as overmerged systems. This might be the reason for the two points in Figure 17 at an overmerging rate of unity (black and red points at $M \sim 10^{14.75}$ and $M \sim 10^{15.25} M_{\odot}$). Both of these points correspond to one single cluster that is overmerged for all detection algorithms.

The fragmentation and over-merging rates are limited in terms of quality assessment because they are largely dependent on the matching procedure. Nevertheless, they provide complementary information and have an impact on the scatter in the scaling relations, depending on how the fragments are distributed in terms of richness. In Figure 13, we observe outliers that are likely related to over-merging, above the mean relation, and fragmentation, below the mean relation. Fragmentation and over-merging processes have been investigated recently on BAHAMAS hydrodynamical simulations (McCarthy et al. 2017) in the context of a friend-of-friend cluster finder algorithm, showing in particular the impact of fragmentation on the scaling relations (Jakobs et al. 2018).

6. Discussion

6.1. Global performance, algorithm selection and comparison to previous work

Among the six codes that have been tested in the final CFC, four of them reach a mean completeness of 80% for masses larger than $10^{14} M_{\odot}$, with high values of purity as shown in Table 1. AMICO and PZWav demonstrate excellent performance at high redshift in terms of completeness and purity (as a function of ranking), which is of particular importance for cosmological purposes. Based on these results, as detailed in Section 5, the algorithm presenting the best overall performance, in the context of this work, is the AMICO code. AMICO was therefore selected to be implemented in the *Euclid* pipeline.

Because of the current uncertainties in the physical processes that drive cluster formation at high redshift, which are one of the main targets of *Euclid*, it is important for the detection algorithms to be robust with respect to the underlying cluster properties. In Table 3, we provide the complementarity of the cluster finder catalogs with respect to AMICO. This shows the gain achieved on the completeness when using AMICO plus another code. This helps in particular to increase the completeness at the high mass and high redshift. The most efficient codes in terms of complementarity to AMICO are WaZP, sFoF and PZWav, with relative performance varying according to the mass and redshift ranges considered. Among these three codes, PZWav has the best performance in terms of completeness and purity at high redshifts (see Figures 11 and 12). This is the regime where *Euclid* is expected to have the most impact on cluster detection with respect to optical ground based and X-ray surveys. PZWav was thus also selected to be implemented in the *Euclid* cluster detection pipeline. The detailed implementation of the two codes and their joint utilization is under investigation and we leave this point for future publications.

AMICO and PZWav detected around 7700 and 7100 clusters, respectively, in a 300 deg^2 region with a purity of 80% for $M > 10^{14} M_{\odot}$ and up to $z_{\text{true}} = 2$. Assuming that these numbers scale with the *Euclid* survey area of 15000 deg^2 , we expect a total number of clusters of $\sim 3.5 \times 10^5$ in the same mass and redshift range. At redshifts between 1 and 2, this number becomes $\sim 10^5$. These numbers of clusters are in reasonable agreement with those obtained by Sartoris et al. (2016) (i.e., $\sim 2 \times 10^5$ at all redshifts and $\sim 4 \times 10^4$ at $z_{\text{true}} \geq 1$, for $S/N \gtrsim 5$). At a $S/N \gtrsim 3$, these numbers are larger by a factor of 10^4 .

The *Euclid* cluster selection function was also investigated using the same mock as the one used in the present paper

⁴ Here, the quoted S/N is defined as a number of galaxies within R_{500} normalized by the rms of the field counts within the same radius, see Sartoris et al. (2016) for more details.

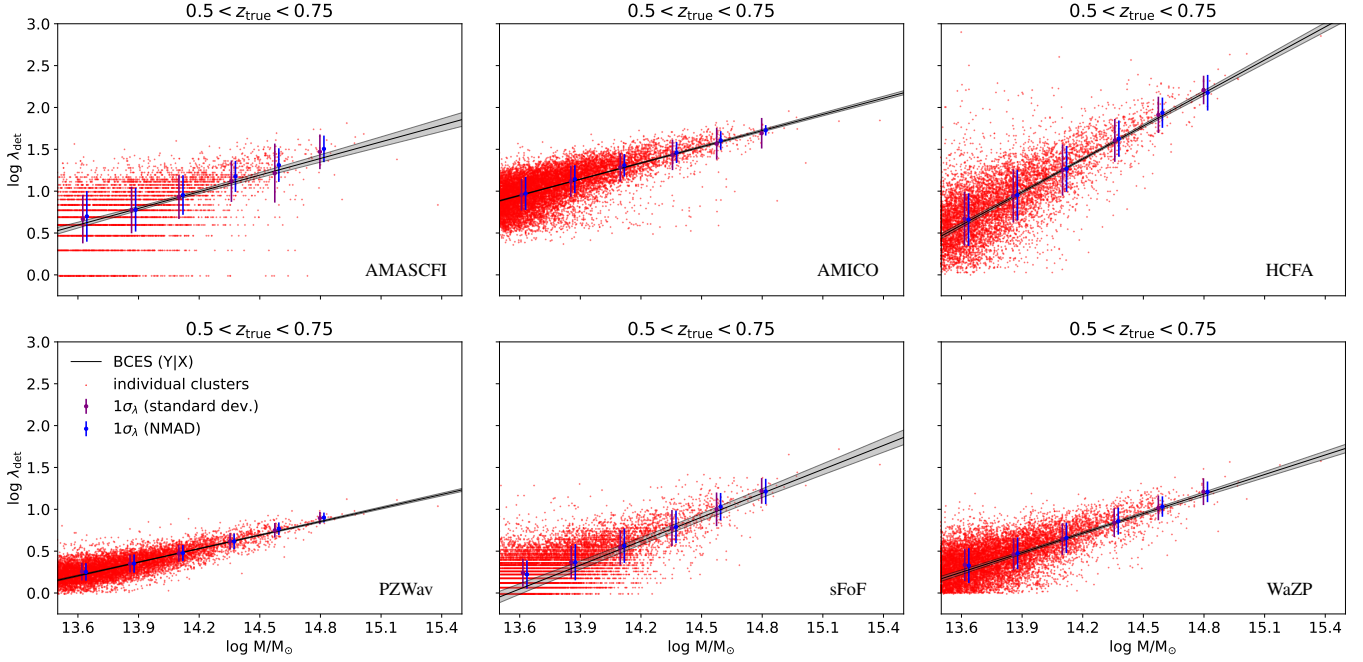


Figure 13: Example of the mass-richness scaling, for redshifts in the range $z_{\text{true}} = [0.5, 0.75]$ for all the cluster finders. The legend is the same as in Figure 3. To improve statistics, the full detection catalogs, corresponding to true detections, were used to produce the figure. The original richness provided by the finders was renormalized for better comparison. Note that the richness provided by AMASCFI and sFoF are discrete, while it is continuous for the other codes.

Table 3: Complementarity of the final CFC detection algorithm with respect to the AMICO results. Two redshift ranges are considered, $[0, 1]$, and $[1, 2]$, both for masses larger than $10^{14} M_{\odot}$. The two numbers correspond to the common fraction of clusters, with respect to AMICO, and the fraction of extra detections, with respect to the total. For example, adding PZWAV to AMICO, in the second redshift bin, would add 5.8% extra clusters, while they have already 81.6% clusters in common.

Finder	$z = [0, 1], M > 10^{14} M_{\odot}$	$z = [1, 2], M > 10^{14} M_{\odot}$
–	(common fraction, fraction of extra detections)	
AMASCFI	(71.2, 2.7)%	(60.7, 4.6)%
HCFA	(59.0, 2.2)%	(53.6, 4.5)%
PZWav	(89.3, 3.2)%	(81.6, 5.8)%
sFoF	(60.3, 2.0)%	(59.5, 3.4)%
WaZP	(90.1, 4.2)%	(78.9, 5.5)%

(Ascaso et al. 2017). Galaxy clusters were detected using the Bayesian Cluster Finder (BCF, Ascaso et al. 2012, 2014, 2015). The results by Ascaso et al. (2017) are in broad agreement with those presented in this paper, but are based on a different methodology. Nonetheless, the exact shape of the selection function is different, as it also depends on the algorithm considered (almost flat at $z_{\text{true}} < 1$ and increases at higher redshifts in Ascaso et al. 2017).

6.2. Representativity of the mock and limitations

The mock used was constructed to be representative of expected *Euclid* data. However, it was originally designed for large-scale galaxy clustering studies, and not for galaxy cluster studies. It is known to present limitations, as discussed in Section 3. For instance, clusters are more concentrated than expected from obser-

vations and the density profiles of clusters are truncated, which could affect the detection performance. The LF also differs from the expectations of passive evolution, depending on the redshift regime we are interested in. The differences that we observe among the different detection algorithms are based on different information and hypotheses regarding the cluster properties. These differences reflect the intrinsic performance of the codes, their sensitivity to the underlying structural properties of galaxy clusters and their sensitivity to photometric errors. The accuracy to which the *Euclid* selection function will be determined and how systematic effects impact on the derived cosmological constraints will strongly depend on our ability to understand and model the properties of *Euclid* clusters. The work presented in this paper gives the current status of the performance of cluster detection within *Euclid*. This assessment is limited by the simulation used, which is not necessarily fully representative of the true Universe, especially since the properties of distant clusters remain poorly known to date. Nonetheless, despite the fact that the mock may not be fully representative of the true Universe, the relative performance of the algorithms is expected to be fairly stable given their behavior during previous challenges on other mocks (see the Appendix for further details).

In addition to the cluster properties, we note that the cosmological parameters assumed in the simulation differ from current constraints, in particular $\Omega_m = 0.25$ and $\sigma_8 = 0.9$ (versus 0.3156 ± 0.0091 and 0.831 ± 0.013 , Planck Collaboration et al. 2016c). This is expected to lead to fewer projection effects with other clusters along the line-of-sight and to increase the overall number of clusters.

Finally, we note that the photometric redshifts were computed using *Euclid* *YJH* bands together with *grizY* bands from assumed ground-based observations. We stress that the addition of *u*-band data, in a more optimistic case, is expected to significantly improve the photometric redshift quality (Ascaso et al.

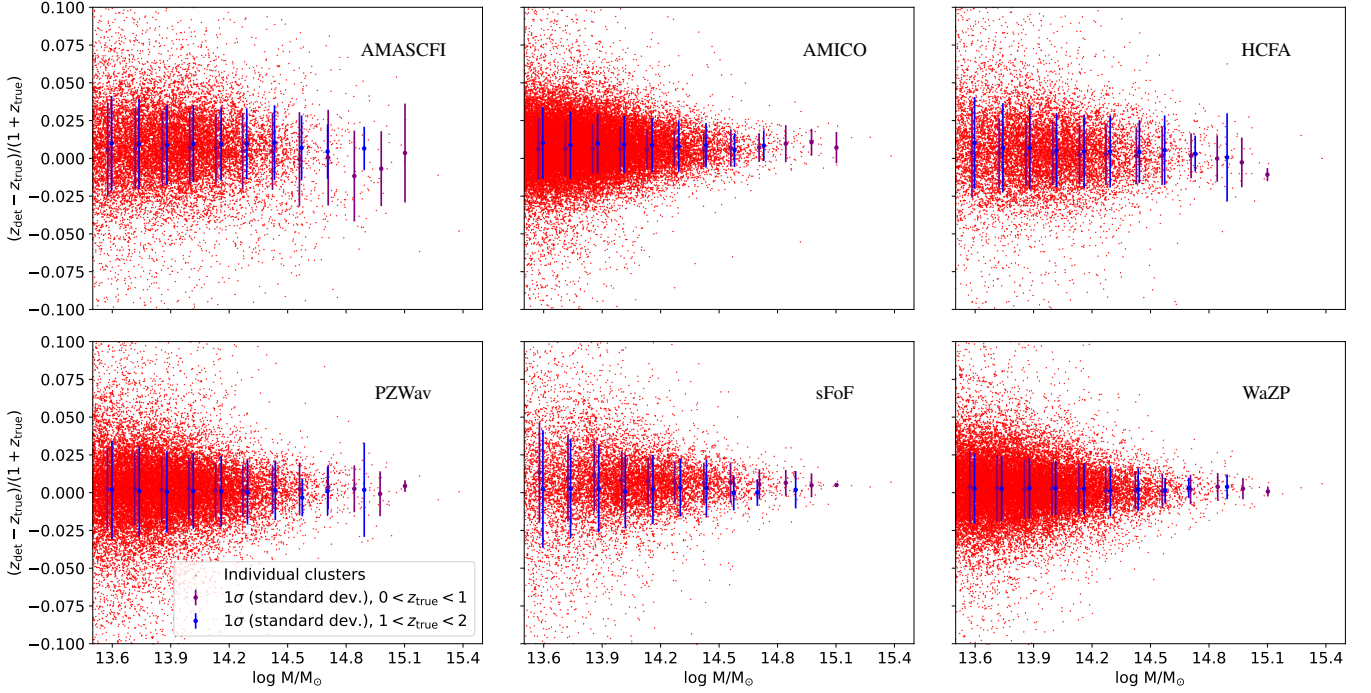


Figure 14: Difference between the redshift associated to a detected cluster and the mock cluster redshift from the simulation, as a function of mock cluster mass, for all cluster finders. The red points show the individual clusters and the error bars provide the standard deviation of the distribution as a function of mass (in bins of 0.14 dex) for two redshift bins. Each catalog of detections has been trimmed to the most reliable detections ensuring a mean purity of 80% in the range $M > 10^{14} M_{\odot}$ and $z_{\text{true}} < 2$.

2015), which in turn can improve the performance of the cluster finders. *Euclid* ground-based complementary observations are currently under way and, combined with the large surveys expected to be released at the time *Euclid* data will be available, will improve the precision and robustness of photometric redshifts. Furthermore, our analysis is based on the *Euclid* wide photometric survey. Using *Euclid* spectroscopic data will provide detection of the $H\alpha$ line down to a flux limit of $2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ ($5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ for the deep survey) over the redshift interval $0.9 < z < 1.8$, so that we expect to further improve the detections of clusters in the high redshift regime (e.g., improve the purity by identifying wrong galaxy associations in redshifts).

In this work, we restricted the mock cluster catalog to masses of $M > 10^{13.25} M_{\odot}$, corresponding on average to $M_{200} > 10^{13.15} M_{\odot}$, and even less for M_{500} , depending on the cluster concentration. Given this choice, lower mass objects that can be detected by cluster finders appear as impurities. Therefore, the established performance presented here is conservative. By using different mass cuts for the catalog, we have verified that these objects do not significantly affect our findings.

Given the limited mock area and since clusters are extended objects, clusters for which the central galaxy was out of the footprint were not included in the mock cluster catalog. These objects can also show up as impurities in the detections, depending on how the cluster finders deal with edges. However, by using a catalog made from the barycenter of cluster member galaxies (i.e., including all the galaxies within the footprint), we have observed that edge effects are subdominant.

In this paper, our baseline choice for comparing the performance was to use a purity threshold of 80% for the different cluster finder catalogs. Changing this limit leads to an overall shift in the accessible mass limit at a given completeness that is

similar for all algorithms. For instance, it can be seen on average in Figure 8, for the range $z_{\text{true}} = [0, 2]$ and $M > 10^{14} M_{\odot}$, that the completeness is reduced by a factor of typically ~ 1.4 when requiring a purity of 90%. When focussing on the high mass range, $M > 10^{14.5} M_{\odot}$, the completeness is much less affected with an overall mean reduction by a factor of ~ 1.03 . The best tradeoff between purity and completeness and the best range of mass that should be considered will be further addressed in future work.

7. Summary and conclusions

In this paper, we have presented the methodology and the results of the final *Euclid* Cluster Finder Challenge. This activity was organized to estimate the performance of galaxy cluster detection algorithms within *Euclid* and to select the cluster finders to be implemented in the *Euclid* pipeline. A total of six algorithms, based on various assumptions and techniques, were considered. They were applied to a galaxy mock that is believed to provide a fairly good representation of expected *Euclid* data, albeit being significantly smaller in terms of sky coverage. We estimated the performance of the different algorithms by matching the detected clusters to the known mock clusters. We observe that the mass-dependent bias on completeness, due to the matching procedure, is estimated to be below 3% over the mass range considered. If unaccounted for, such biases may affect the derived mass-observable scaling relations, as well as the normalization of cluster counts. Several other methods to estimate the selection function are being developed within the consortium and they should allow us to mitigate such effects.

All six algorithms performed well, three of them reaching a completeness and purity higher than 80%, down to masses of $10^{14} M_{\odot}$ and up to redshift of 2. Among the competing cluster

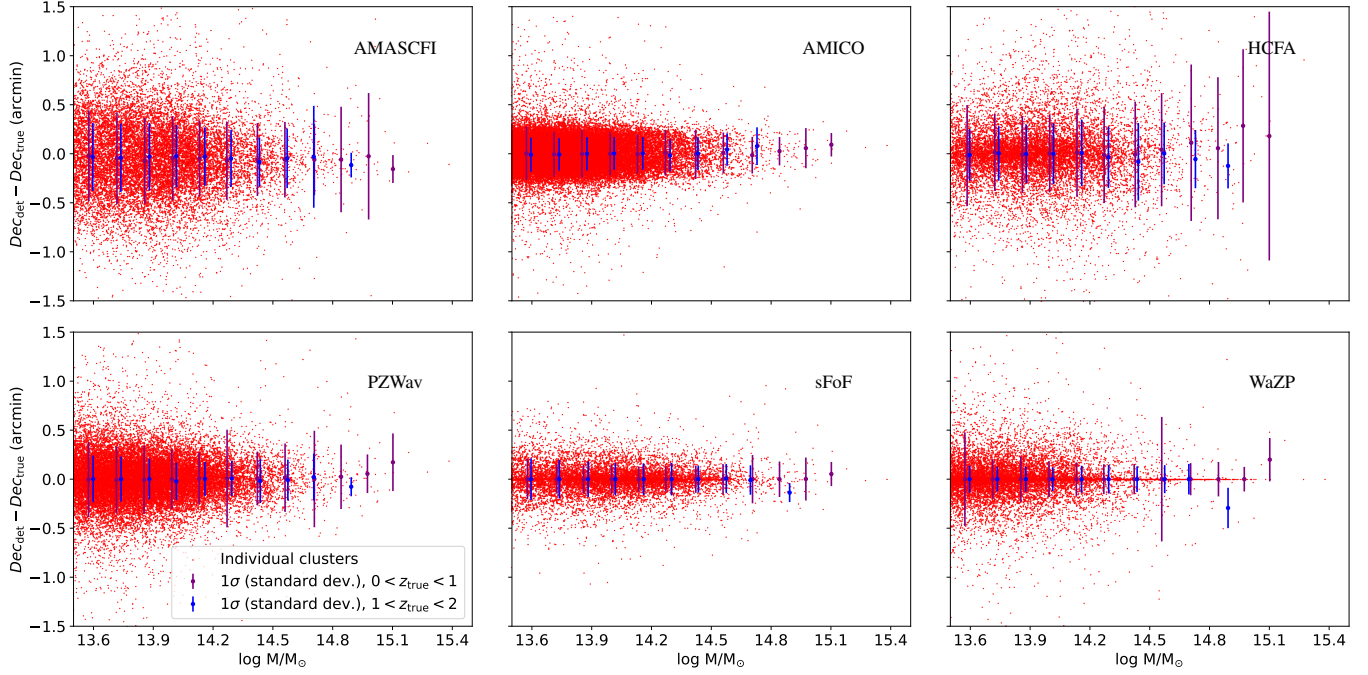


Figure 15: Difference between the declination associated to a detected cluster and the mock cluster declination from the simulation, as a function of mock cluster mass, for all cluster finders. The red points show the individual clusters and the error bars provide the standard deviation of the distribution as a function of mass (in bins of 0.14 dex width) for two redshift bins. Each catalog of detections has been trimmed to the most reliable detections insuring a mean purity of 80% in the range $M > 10^{14} M_{\odot}$ and $z_{\text{true}} < 2$.

finders, the AMICO and PZWav codes were selected to be implemented in the *Euclid* cluster pipeline.

This work was based on the application of cluster finders to a galaxy mock, which was characterized, and found to show some limitations in the context of galaxy cluster detection. The physical processes at play in distant clusters (the ones which should provide most of the cosmological power for *Euclid*, Sartoris et al. 2016) are not yet fully understood (e.g., space densities, structural properties, star formation rate). Thus, improving the robustness of our results would require a better modeling of the structural properties of these objects, to be implemented in the mock. Indeed, any mismodeling in the mock could affect the absolute estimate of performance of a given algorithm. In this context, data from the deep surveys and from other surveys of extremely distant clusters will help to characterize the high-redshift tail of *Euclid* clusters and improve the performance of cluster finder algorithms as well as the precision on its absolute determination.

We note that the performance of the detection codes presented in this paper are those at the time of this selection. The codes continue to be developed and optimized, as well as more representative simulations of the *Euclid* survey become available. Our knowledge of cluster properties at high redshift is expected to significantly improve in the coming years with the advent of new facilities (e.g., *JWST*) and may impact cluster detection methodologies. New methodologies may also become available in coming years. For these reasons, integration of the codes into the *Euclid* pipeline is being configured in a flexible way to allow for the possibility of updating and adding codes, given valid scientific motivations.

The results reported in this paper show that, with current cosmological parameters, *Euclid* has the potential of detecting an unprecedented number of galaxy clusters ($> 10^5$), up to redshift

2 and over more than two orders of magnitude in mass (down to $10^{13.5} M_{\odot}$). We note, however, that predictions of number counts are subject to modeling uncertainty; in particular, as we have commented above, our simulations do not represent the currently favored cosmological parameters (Planck Collaboration et al. 2016b), with larger σ_8 and lower matter density, and do not reproduce certain known cluster properties. Nonetheless, the analytical selection function assumed in Sartoris et al. (2016) and the one obtained in Ascaso et al. (2017) are in line with our findings, such that the accuracy of the Sartoris et al. (2016) forecast is strengthened by the present work. The exquisite leverage provided by *Euclid*, in terms of mass and redshifts, thus enable very competitive constraints on cosmological parameters from cluster number counts. As pointed in Sartoris et al. (2016), the main challenges will be calibrating the mass-richness relation for the sample and characterizing the selection function in a mass and redshift regime that remains relatively unexplored with current observations. An in-depth investigation of the systematic effects related to the mass determination will be possible thanks to the wide mass and redshift range accessible with *Euclid*, its internal mass calibration available from WL and velocity dispersion, and multi-wavelength synergies available at the time *Euclid* is operating.

While this paper presents an estimate of the expected *Euclid* performance in terms of galaxy cluster detection, many developments are still ongoing within the consortium. In particular, we note that activities dedicated to the assessment of the cluster galaxy membership, the characteristic radius of the detected clusters, and the definition of the richness are being pursued. Improvements in the purity will also be achieved by using spectroscopic data, while an internal calibration of the mass-richness relation will benefit from masses estimated from stacking weak-lensing signal (see Köhlinger et al. 2015, for the assessment of

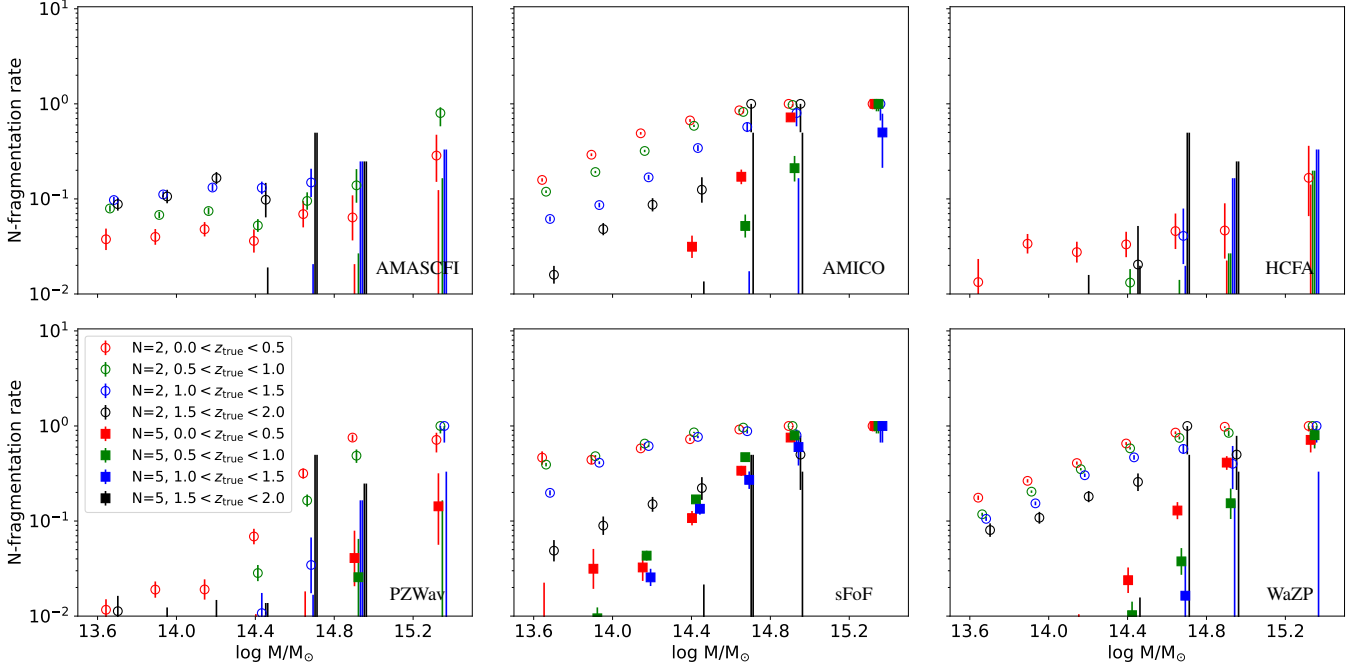


Figure 16: Fraction of fragmented mock clusters as a function of mass, for different redshift bins, for six detection algorithms. The figure shows the fragmentation for a number of fragments larger than one (two or more) and four (five or more), as given in the legend. The symbols that are not visible in the figure correspond to small fragmentation rate, below the figure range. Error bars are computed as in Figure 11. Each catalog of detections has been trimmed to the most reliable detections insuring a mean purity of 80% in the range $M > 10^{14} M_{\odot}$ and $z_{\text{true}} < 2$.

the performance in the *Euclid* context). Therefore, significant progress is possible within the coming years. In addition, the work presented here will be reproduced using different simulations where clusters may present different properties, following improvements in our understanding of cluster formation and evolution, particularly at high redshift. Similarly, the choice of the photometric redshift codes may have an impact on the cluster detection, which will be quantified by testing the available codes.

Acknowledgements. We are thankful to the anonymous referee for useful comments that helped improve the quality of the paper. This work is part of the ongoing effort dedicated to the scientific preparation of the *Euclid* mission and we are grateful to the *Euclid* consortium. The *Euclid* Consortium acknowledges the European Space Agency and the support of a number of agencies and institutes that have supported the development of *Euclid*. A detailed complete list is available on the *Euclid* web site (<http://www.euclid-ec.org>). In particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian *Euclid* Consortium, the Centre National d'Etudes Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Economía y Competitividad, the National Aeronautics and Space Administration, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Center, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. This work is based on simulation products created for the *Euclid* Consortium. They were created on the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1 and ST/K00087X/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. Rémi Adam acknowledges support from Spanish Ministerio de Economía y Competitividad (MINECO) through grant number AYA2015-66211-C2-2. Rémi Adam acknowledges fundings from the CNES post-doctoral fellowship program. Christophe Benoist, Alberto Cappi, Sophie Maurogordato, Marina Ricci, Pier-Francesco Rocci and Martin Vannier acknowledge funding from the CNES program (CNES/INSU). Pier-Francesco Rocci acknowledges funding from a CNES grant. Fabio Bellagamba thanks the support from

the grants ASI n.I/023/12/0 “Attività relative alla fase B2/C per la missione *Euclid*”. Fabio Bellagamba and Stefano Andreon thank the support PRIN MIUR 2015 “Cosmology and Fundamental Physics: Illuminating the Dark Universe with *Euclid*”. Matteo Maturi was supported by the SFB-Transregio TR33 “The Dark Universe”. Anastasio Díaz-Sánchez acknowledges support from project ESP2015-69020-C2-1-R (MINECO). Anthony Gonzalez acknowledges support from NASA ROSES grant 12-EUCLID12-0004. Florence Durret acknowledges long term funding from CNES. This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013), in addition to NumPy (van der Walt et al. 2011), SciPy (Jones et al. 2001) and IPython (Pérez & Granger 2007). Figures were generated using Matplotlib (Hunter 2007).

References

- Abbott, T. M. C. et al. 2018, *ApJS*, 239, 18
- Abell, G. O. 1958, *ApJS*, 3, 211
- Adam, R. et al. 2015, *A&A*, 576, A12
- Adami, C. et al. 2010, *A&A*, 509, A81
- Adami, C., & Mazure, A. 1999, *A&AS*, 134, 393
- Adami, C., Ulmer, M. P., Romer, A. K., Nichol, R. C., Holden, B. P., & Pildis, R. A. 2000, *ApJS*, 131, 391
- Akritis, M. G., & Bershady, M. A. 1996, *ApJ*, 470, 706
- Alberts, S. et al. 2016, *ApJ*, 825, 72
- Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, 49, 409
- Amendola, L. et al. 2013, *Living Reviews in Relativity*, 16, 6
- Andreon, S. 2015, *A&A*, 582, A100
- Andreon, S., Maughan, B., Trinchieri, G., & Kurk, J. 2009, *A&A*, 507, 147
- Andreon, S., Punzi, G., & Grado, A. 2005, *MNRAS*, 360, 727
- Annuziatella, M. et al. 2014, *A&A*, 571, A80
- Ascaso, B., Mei, S., Bartlett, J. G., & Benítez, N. 2017, *MNRAS*, 464, 2270
- Ascaso, B., Mei, S., & Benítez, N. 2015, *MNRAS*, 453, 2515
- Ascaso, B., Wittman, D., & Benítez, N. 2012, *MNRAS*, 420, 1167
- Ascaso, B., Wittman, D., & Dawson, W. 2014, *MNRAS*, 439, 1980
- Astropy Collaboration et al. 2013, *A&A*, 558, A33
- Aurenhammer, F., Klein, R., & Lee, D.-T. 2013, *Voronoi Diagrams and Delaunay Triangulations*, 1st edn. (River Edge, NJ, USA: World Scientific Publishing Co., Inc.)
- Barkhouse, W. A., Yee, H. K. C., & López-Cruz, O. 2007, *ApJ*, 671, 1471

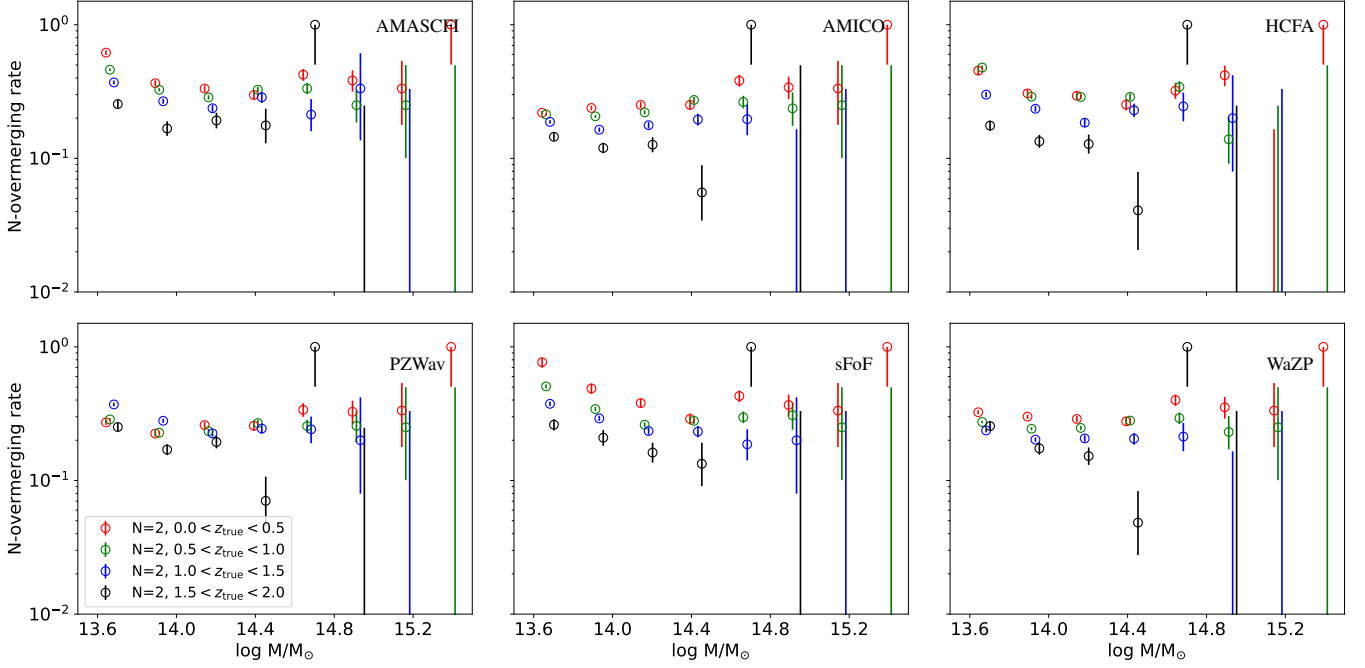


Figure 17: Fraction of overmerged detected clusters (two or more mock clusters associated with a detection), as a function of mass, for different redshift bins, for six detection algorithms. Error bars are computed as in Figure 11. Each catalog of detections has been trimmed to the most reliable detections insuring a mean purity of 80% in the range $M > 10^{14} M_{\odot}$ and $z_{\text{true}} < 2$.

- Bellagamba, F., Maturi, M., Hamana, T., Meneghetti, M., Miyazaki, S., & Moscardini, L. 2011, *MNRAS*, 413, 1145
- Bellagamba, F., Roncarelli, M., Maturi, M., & Moscardini, L. 2018, *MNRAS*, 473, 5221
- Bellagamba, F. et al. 2019, *MNRAS*, 484, 1598
- Benítez, N. 2000, *ApJ*, 536, 571
- Benitez, N. et al. 2014, *ArXiv e-prints*:1403.5237
- Benítez, N. et al. 2004, *ApJS*, 150, 1
- Benoist, C. 2014, in *Building the Euclid Cluster Survey - Scientific Program*, proceedings of a conference held July 6-11 2014 at the Sexten Center for Astrophysics., 8
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Bleem, L. E. et al. 2015, *ApJS*, 216, 27
- Böhringer, H., Chon, G., & Collins, C. A. 2014, *A&A*, 570, A31
- Böhringer, H. et al. 2001, *A&A*, 369, 826
- Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, *A&A*, 363, 476
- Boselli, A. et al. 1997, *A&A*, 324, L13
- Botzler, C. S., Snigula, J., Bender, R., & Hopp, U. 2004, *MNRAS*, 349, 425
- Brodwin, M. et al. 2012, *ApJ*, 753, 162
- . 2013, *ApJ*, 779, 138
- Budzynski, J. M., Kopev, S. E., McCarthy, I. G., McGee, S. L., & Belokurov, V. 2012, *MNRAS*, 423, 104
- Cannon, R. et al. 2006, *MNRAS*, 372, 425
- Cappellari, M., & Copin, Y. 2003, *MNRAS*, 342, 345
- Carlberg, R. G. et al. 1997, *ApJ*, 485, L13
- Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, *MNRAS*, 447, 646
- Castignani, G., & Benoist, C. 2016, *A&A*, 595, A111
- Cava, A. et al. 2017, *A&A*, 606, A108
- Ciliegi, P. et al. 2005, *A&A*, 441, 879
- Coe, D., Benítez, N., Sánchez, S. F., Jee, M., Bouwens, R., & Ford, H. 2006, *AJ*, 132, 926
- Collister, A. A., & Lahav, O. 2005, *MNRAS*, 361, 415
- Costanzi, M. et al. 2019, *MNRAS*, 482, 490
- Cucciati, O. et al. 2010, *A&A*, 520, A42
- Dalton, G. B., Maddox, S. J., Sutherland, W. J., & Efstathiou, G. 1997, *MNRAS*, 289, 263
- de Haan, T. et al. 2016, *ApJ*, 832, 95
- De Lucia, G., Weinmann, S., Poggianti, B. M., Aragón-Salamanca, A., & Zaritsky, D. 2012, *MNRAS*, 423, 1277
- De Propriis, R. 2017, *MNRAS*, 465, 4035
- de Propriis, R., Eisenhardt, P. R., Stanford, S. A., & Dickinson, M. 1998, *ApJ*, 503, L45
- de Propriis, R., & Pritchet, C. J. 1998, *AJ*, 116, 1118
- de Propriis, R., Stanford, S. A., Eisenhardt, P. R., Dickinson, M., & Elston, R. 1999, *AJ*, 118, 719
- Diehl, S., & Statler, T. S. 2006, *MNRAS*, 368, 497
- Dietrich, J. P. et al. 2014, *MNRAS*, 443, 1713
- Driver, S. P., Phillipps, S., Davies, J. I., Morgan, I., & Disney, M. J. 1994, *MNRAS*, 268, 393
- Durret, F. et al. 2015, *A&A*, 578, A79
- . 2011, *A&A*, 535, A65
- Eisenhardt, P. R. M. et al. 2008, *ApJ*, 684, 905
- Ellis, S. C., & Jones, L. R. 2004, *MNRAS*, 348, 165
- Elston, R. J. et al. 2006, *ApJ*, 639, 816
- Farrens, S., Abdalla, F. B., Cypriano, E. S., Sabiu, C., & Blake, C. 2011, *MNRAS*, 417, 1402
- Fioc, M., & Rocca-Volmerange, B. 1997, *A&A*, 326, 950
- Gal, R. R. 2006, *ArXiv e-prints*:astro-ph/0601195
- Gal, R. R., de Carvalho, R. R., Lopes, P. A. A., Djorgovski, S. G., Brunner, R. J., Mahabal, A., & Odewahn, S. C. 2003, *AJ*, 125, 2064
- Gavazzi, R., & Soucail, G. 2007, *A&A*, 462, 459
- George, M. R. et al. 2011, *ApJ*, 742, 125
- Gilbank, D. G., Gladders, M. D., Yee, H. K. C., & Hsieh, B. C. 2011, *AJ*, 141, 94
- Gladders, M. D., & Yee, H. K. C. 2000, *AJ*, 120, 2148
- Gonzalez, A. 2014, in *Building the Euclid Cluster Survey - Scientific Program*, proceedings of a conference held July 6-11 2014 at the Sexten Center for Astrophysics., 7
- Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., Lagos, C. D. P., Helly, J., Campbell, D. J. R., & Mitchell, P. D. 2014, *MNRAS*, 439, 264
- Goto, T. et al. 2002, *AJ*, 123, 1807
- Hasselfield, M. et al. 2013, *J. Cosmology Astropart. Phys.*, 7, 8
- Henriques, B. M. B., White, S. D. M., Lemson, G., Thomas, P. A., Guo, Q., Marleau, G.-D., & Overzier, R. A. 2012, *MNRAS*, 421, 2904
- Huchra, J. P., & Geller, M. J. 1982, *ApJ*, 257, 423
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Icke, V., & van de Weygaert, R. 1991, *QJRAS*, 32, 85
- Ilbert, O. et al. 2006, *A&A*, 457, 841
- Jakobs, A. et al. 2018, *MNRAS*, 480, 3338
- Jeffrey, N. et al. 2018, *MNRAS*, 479, 2871
- Jiang, L., Helly, J. C., Cole, S., & Frenk, C. S. 2014, *MNRAS*, 440, 2115

- Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python, <http://www.scipy.org/>
- Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R., & Xu, G. 1999, *ApJ*, 517, 78
- Kim, R. S. J. et al. 2002, *AJ*, 123, 20
- Knobel, C. et al. 2009, *ApJ*, 697, 1842
- Koester, B. P. et al. 2007, *ApJ*, 660, 239
- Köhlinger, F., Hoekstra, H., & Eriksen, M. 2015, *MNRAS*, 453, 3107
- Lagos, C. d. P., Bayet, E., Baugh, C. M., Lacey, C. G., Bell, T. A., Fanidakis, N., & Geach, J. E. 2012, *MNRAS*, 426, 2142
- Laureijs, R. et al. 2011, *ArXiv e-prints*:1110.3193
- Licitra, R., Mei, S., Raichoor, A., Erben, T., & Hildebrandt, H. 2016a, *MNRAS*, 455, 3020
- Licitra, R. et al. 2016b, *ApJ*, 829, 44
- Lin, Y.-T., Mohr, J. J., Gonzalez, A. H., & Stanford, S. A. 2006, *ApJ*, 650, L99
- Lin, Y.-T., Mohr, J. J., & Stanford, S. A. 2004, *ApJ*, 610, 745
- LSST Science Collaboration et al. 2009, *ArXiv e-prints*:0912.0201
- Majumdar, S., & Mohr, J. J. 2004, *ApJ*, 613, 41
- Mamon, G. A., Biviano, A., & Murante, G. 2010, *A&A*, 520, A30
- Mana, A., Giannantonio, T., Weller, J., Hoyle, B., Hütsi, G., & Sartoris, B. 2013, *MNRAS*, 434, 684
- Mannucci, F., Basile, F., Poggianti, B. M., Cimatti, A., Daddi, E., Pozzetti, L., & Vanzì, L. 2001, *MNRAS*, 326, 745
- Mantz, A. B. et al. 2015, *MNRAS*, 446, 2205
- Marinoni, C., Davis, M., Newman, J. A., & Coil, A. L. 2002, *ApJ*, 580, 122
- Maturi, M., Bellagamba, F., Radovich, M., Roncarelli, M., Sereno, M., Moscardini, L., Bardelli, S., & Puddu, E. 2019, *MNRAS*, 485, 498
- Maturi, M., Meneghetti, M., Bartelmann, M., Dolag, K., & Moscardini, L. 2005, *A&A*, 442, 851
- Mazure, A. et al. 2007, *A&A*, 467, 49
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, *MNRAS*, 465, 2936
- Mei, S. et al. 2009, *ApJ*, 690, 42
- . 2015, *ApJ*, 804, 117
- Merson, A. I. et al. 2013, *MNRAS*, 429, 556
- Mohr, J. J. et al. 2012, in *Proc. SPIE*, Vol. 8451, Software and Cyberinfrastructure for Astronomy II, 84510D
- Molino, A. et al. 2017, *MNRAS*, 470, 95
- Muzzin, A., Yee, H. K. C., Hall, P. B., Ellingson, E., & Lin, H. 2007, *ApJ*, 659, 1106
- Nakata, F. et al. 2001, *PASJ*, 53, 1139
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Neyrinck, M. C. 2008, *MNRAS*, 386, 2101
- Noiro, G. et al. 2016, *ApJ*, 830, 90
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. 2000, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd edn., Series in Probability and Statistics (John Wiley and Sons, Inc.)
- Old, L. et al. 2015, *MNRAS*, 449, 1897
- Olsen, L. F. et al. 2007, *A&A*, 461, 81
- Pacaud, F. et al. 2016, *A&A*, 592, A2
- Parroni, C. et al. 2017, *ApJ*, 848, 114
- Pérez, F., & Granger, B. E. 2007, *Computing in Science and Engineering*, 9, 21
- Pierre, M. 1990, *A&A*, 229, 7
- Pierre, M. et al. 2016, *A&A*, 592, A1
- Planck Collaboration et al. 2014, *A&A*, 571, A20
- . 2016a, *A&A*, 594, A27
- . 2016b, *A&A*, 594, A13
- . 2016c, *A&A*, 594, A24
- Platen, E., van de Weygaert, R., & Jones, B. J. T. 2007, *MNRAS*, 380, 551
- Popesso, P., Böhringer, H., Romaniello, M., & Voges, W. 2005, *A&A*, 433, 415
- Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P., & Christensen, J. A. 1996, *AJ*, 111, 615
- Radovich, M. et al. 2017, *A&A*, 598, A107
- Raichoor, A., & Andreon, S. 2012, *A&A*, 543, A19
- Ramella, M., Boschin, W., Fadda, D., & Nonino, M. 2001, *A&A*, 368, 776
- Ramella, M., Boschin, W., Geller, M. J., Mahdavi, A., & Rines, K. 2004, *AJ*, 128, 2022
- Rettura, A. et al. 2014, *ApJ*, 797, 109
- Ricci, M. et al. 2018, *A&A*, 620, A13
- Rozo, E. et al. 2010, *ApJ*, 708, 645
- Rykoff, E. S. et al. 2014, *ApJ*, 785, 104
- Saro, A. et al. 2015, *MNRAS*, 454, 2305
- Sarron, F., Martinet, N., Durret, F., & Adami, C. 2018, *A&A*, 613, A67
- Sartoris, B. et al. 2016, *MNRAS*, 459, 1764
- Schechter, P. 1976, *ApJ*, 203, 297
- Shan, H. et al. 2012, *ApJ*, 748, 56
- Spergel, D. et al. 2015, *ArXiv e-prints*:1503.03757
- Springel, V. et al. 2005, *Nature*, 435, 629
- Sridhar, S., Maurogordato, S., Benoist, C., Cappi, A., & Marulli, F. 2017, *A&A*, 600, A32
- Stanford, S. A. et al. 2012, *ApJ*, 753, 164
- Starck, J.-L., Murtagh, F. D., & Bijaoui, A. 1998, *Image Processing and Data Analysis*, 297
- Strazzullo, V. et al. 2016, *ApJ*, 833, L20
- . 2010, *A&A*, 524, A17
- . 2006, *A&A*, 450, 909
- Sunyaev, R. A., & Zel'dovich, Y. B. 1972, *Astrophys. Space Phys. Res.*, 4, 173
- Toft, S., Mainieri, V., Rosati, P., Lidman, C., Demarco, R., Nonino, M., & Stanford, S. A. 2004, *A&A*, 422, 29
- Trevisan, M., & Mamon, G. A. 2017, *MNRAS*, 471, 2022
- Trevisan, M., Mamon, G. A., & Stalder, D. H. 2017, *MNRAS*, 471, L47
- van der Burg, R. F. J., Hoekstra, H., Muzzin, A., Sifón, C., Balogh, M. L., & McGee, S. L. 2015, *A&A*, 577, A19
- van der Burg, R. F. J., Muzzin, A., Hoekstra, H., Wilson, G., Lidman, C., & Yee, H. K. C. 2014, *A&A*, 561, A79
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, 13, 22
- Veropalumbo, A., Marulli, F., Moscardini, L., Moresco, M., & Cimatti, A. 2014, *MNRAS*, 442, 3275
- Vikhlinin, A. et al. 2009, *ApJ*, 692, 1060
- Wen, Z. L., Han, J. L., & Liu, F. S. 2012, *ApJS*, 199, 34
- Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, *MNRAS*, 432, 336
- Wylezalek, D. et al. 2013, *ApJ*, 769, 79
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2008, *ApJ*, 676, 248

Appendix A: Summary of the previous challenges

In preparation for the final cluster finder challenge described in this paper, three other challenges took place between 2013 and 2017, organized within the Work-Package Clusters of Galaxies (Implementation) of the *Euclid* Science Ground Segment Organisation Unit Level3. The aim of these preparatory challenges was to set up and refine the procedure by which to assess the relative performance of the different cluster finder algorithms. Eight different codes were tested on two *Euclid* survey-like mock catalogs, where the positions of the massive halos to be detected by the codes were unknown to the participants to the challenge.

The analysis of these two challenges raised several difficulties that had to be overcome. In particular, the definition of S/N was not common to all cluster finder algorithms, so to compare the relative performance of the algorithms in terms of completeness, we ranked the detections in S/N and selected the first N detections in order of decreasing S/N, separately for each algorithm. To compare the relative performance in terms of purity, we translated the different richness estimates provided by the different algorithms into masses, by fitting a mass-richness relation independently for each algorithm.

To test whether the relative performance of the different codes was dependent on the mock catalog itself, two mocks were considered, one based on semianalytic modeling and one based on halo occupation distribution (HOD). However, we noticed when performing tests on the properties of galaxies in clusters in the different mocks that the cluster number density concentration in the mocks based on HOD simulations was very high (~ 8 times the typical values). This led us to use a mock based on a SAM for the final challenge. Both mocks reproducing the wide and the deep *Euclid* surveys were considered. Initially, mock galaxies were assigned photometric redshifts from a random Gaussian distribution centered on the real galaxy redshift with a dispersion equal to the *Euclid* survey requirement, $\sigma_z = 0.03(1+z)$ or $0.05(1+z)$, and with an outlier fraction of 5% or 10%. At a later stage, photometric redshifts were computed from fits to the galaxy spectral energy distributions (using BPZ by Benítez (2000), Coe et al. (2006), and Hyperz by Bolzonella

et al. (2000)) resulting from the galaxy mock magnitudes with added noise. Galaxy magnitudes in the mocks were rendered closer to real galaxy magnitudes by the use of the PhotReal (Ascaso et al. 2015) algorithm. We found that modifications to the photometric redshift assignment method had little to no impact on the performance of the detection algorithms.

At the end of the third cluster finder challenge, we considered the whole methodology and analysis pipeline to be sufficiently mature to make a final assessment of the relative performance of the different codes. While eight cluster finder codes in total were tested in the three preliminary challenges, only six of them took part in the final challenge described in this paper.

We hereafter give the description of the two codes (RedGOLD and Voronoi) that did not participate in the final challenge for reasons not related to their performance in the earlier CFCs.

A.1. RedGOLD

The RedGOLD cluster detection algorithm (Licitra et al. 2016a,b) was developed as a modified version of color-based detection algorithms such as RedMaPPer (Rykoff et al. 2014), taking into account galaxy morphology and color cuts performed on clusters at high redshifts (Mei et al. 2009, 2015). The algorithm requires as input a catalog with galaxy positions, photometry and photometric redshifts. It selects overdensities of galaxies in a color-color plane, and it is adapted to select both red passive and blue star-forming galaxies. At redshift $z \lesssim 1.5$, where clusters are mostly dominated by a passive galaxy population and show a tight red sequence, the algorithm selects overdensities of red passive galaxies while minimizing contamination from dusty star-forming galaxies. It imposes an NFW profile and calculates cluster detection significance and richness, which is tightly correlated to weak lensing masses (Parroni et al. 2017). In RedGOLD, the S/N is defined as the significance of detections with respect to the background galaxy density, as described in Licitra et al. (2016a). In Licitra et al. (2016a,b), cluster candidate catalogs were obtained for the CFHT-LS and NGVS surveys. When compared to X-ray detected cluster catalogs on these two surveys and the Henriques et al. (2012) mock galaxy catalogs from the Millenium simulation (Springel et al. 2005), RedGOLD was demonstrated to be 80% pure up to $z \sim 1$, and $\sim 100\%$ ($\sim 70\%$) complete at $z \leq 0.6$ ($z \leq 1$) for galaxy clusters with $M > 10^{14} M_{\odot}$. RedGOLD participated in the *Euclid* CFC I and II.

A.2. Voronoi

The Voronoi diagram and its dual, the Delaunay triangulation, have long been known as a very useful and versatile mathematical techniques in a variety of topics - including 3D reconstruction and modeling of objects, visualization of medical datasets, shape analysis and pattern recognition, computer animation, to mention a few (Okabe et al. 2000; Aurenhammer et al. 2013, for wide a review of their applications in computational geometry). In short, a Voronoi tessellation on a two-dimensional (tri-dimensional) distribution of objects is a unique plane (volume) partition into convex cells (polytopes), each of them containing one, and only one, such object. These are the set of points which are closer to that object than to any other. Its main advantage consists in its being completely non-parametric and adaptive, thus very useful in all those cases when one does not desire to define a priori a specific spatial scale for the analysis. The first

order Delaunay neighbors of a chosen object are then those objects in the centers of the cells (polytopes) sharing a wall with the cell (polytope) of the chosen object. Second order Delaunay neighbors are obtained extending one step further out, that is considering also objects that are first order neighbors of objects in the previous list. The Voronoi tessellation has received attention in astrophysics, and among its many usages we may quote those as a tool to reproduce the foamy distribution of galaxies in space (see Pierre 1990; Icke & van de Weygaert 1991), to find voids in galaxy spatial distribution (see Platen et al. 2007; Neyrinck 2008), to bin data to a constant S/N per bin in integral field spectroscopy (Cappellari & Copin 2003) or X-ray imaging data (Diehl & Statler 2006), and to detect clusters both in 2D projected photometric galaxy catalogs (Ramella et al. 2001) and in 3D galaxy redshift catalogs (see Marinoni et al. 2002; Cucciati et al. 2010).

The Voronoi algorithm written for *Euclid* CFC was tailored for cluster detection in photometric redshift space, with a typical error in photometric redshift equal to $\sigma_{z_{\text{phot}}}$. The algorithm works in two steps.

In the first step the sample of galaxies is split into a series of partially overlapping redshift slices. Each slice is defined with thickness equal to $1.5 \sigma_{z_{\text{phot}}}$ and is separated from the adjacent slices by photometric redshift steps of $0.5 \sigma_{z_{\text{phot}}}$ (a galaxy will then usually be present in more than one redshift slice). In each redshift slice the algorithm starts by building the Voronoi tessellation of the RA-Dec galaxy distribution and then computes, for each galaxy, the area covered by including both first and second order Voronoi-Delaunay neighbors, a choice that is driven by the need to minimize the noise in the area estimate. In each redshift slice, all the areas thus estimated are sorted, a fit to their distribution is computed and only galaxies whose area is below 1.5σ the mean value (thus galaxies located in higher than the mean density regions) are kept as cluster seeds. Around each of these seeds we grow outwards, and keep adding first order neighbors of peripheral galaxies. The growth continues as long as the added members satisfy a request on their second order Voronoi-Delaunay neighbors area (should be smaller than a predefined cutoff) and a cutoff on the growth ratio (at least 10 new members added in each subsequent growth step). This procedure is run on all the redshift slices defined in the galaxy catalog.

The second step of the algorithm then merges together (using RA, Dec and z_{phot} information) results from different redshift slices, thus building the final cluster catalog. Each cluster is defined by a center: median of its putative member galaxies RA, Dec and z_{phot} values. The algorithm computes for each cluster its S/N using expected values for a distribution of points with the same surface density as on the actual z_{phot} shell that contains most of its members - if needed during this step further trimming or enlarging of the outer regions can be done, while also a very bland compactness limit ($n_{90}/n_{50} < 2$) is applied, where n_{50}/n_{90} is the number of galaxies included within 50%/90% of the group radius.

The algorithm computes also cluster areas, by summing Voronoi-Delaunay areas of connected galaxies, and observed richness of clusters, after statistical subtraction of expected background galaxy density within the cluster area. This code participated in the *Euclid* CFC I.

¹ Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan, 1, planta 2, E-44001, Teruel, Spain

² Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, France

³ Laboratoire Leprince-Ringuet, Ecole Polytechnique, CNRS/IN2P3, 91128 Palaiseau, France

- ⁴ INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, 34143, Trieste, Italy
- ⁵ Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
- ⁶ Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, F-75014, Paris, France
- ⁷ Dipartimento di Fisica e Astronomia, Università di Bologna, via Gobetti 93/2, I-40129 Bologna, Italy
- ⁸ Istituto Nazionale di Astrofisica (INAF) - Osservatorio di Astrofisica e Scienza dello Spazio (OAS), via Gobetti 93/3, I-40127 Bologna, Italy
- ⁹ Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- ¹⁰ Institut d'Astrophysique de Paris (UMR 7095: CNRS & Sorbonne Université), 98 bis Bd Arago, 75014 Paris, France
- ¹¹ IRFU, CEA, Université Paris-Saclay, F-91191 Gif Sur Yvette, France
- ¹² Department of Astronomy, University of Florida, Bryant Space Science Center, Gainesville, FL 32611, USA
- ¹³ INAF - Osservatorio Astronomico di Brera, via Brera 28, 20122 Milano, via. E. Bianchi 46, 23807 Merate, Italy
- ¹⁴ University of Paris Denis Diderot, University of Paris Sorbonne Cité (PSC), 75205 Paris Cedex 13, France
- ¹⁵ Zentrum für Astronomie, Universität Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany
- ¹⁶ Jet Propulsion Laboratory, Cahill Center for Astronomy & Astrophysics, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California, USA
- ¹⁷ Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- ¹⁸ Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, via Tiepolo 11, I-34131 Trieste - Italy
- ¹⁹ Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark
- ²⁰ Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, F-31400 Toulouse, France
- ²¹ Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland
- ²² Université Paris Diderot, AIM, Sorbonne Paris Cité, CEA, CNRS, F-91191 Gif-sur-Yvette, France
- ²³ APC, AstroParticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, 10 rue Alice Domon et Léonie Duquet, 75205, Paris Cedex 13, France
- ²⁴ Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK
- ²⁵ INFN, Istituto Nazionale di Fisica Nucleare, Trieste, Italy
- ²⁶ Department of Physics and Astronomy, University of Missouri, 5110 Rockhill Road, Kansas City, MO 64110
- ²⁷ Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
- ²⁸ Institut d'Estudis Espacials de Catalunya (IEEC), 08193 Barcelona, Spain
- ²⁹ Collège de France, 11 Place Marcelin Berthelot, 75231 Paris, France
- ³⁰ INFN - Sezione di Bologna, viale Berti-Pichat 6/2, I-40127 Bologna, Italy
- ³¹ Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
- ³² INAF - Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 - Padova, Italy
- ³³ Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- ³⁴ Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany
- ³⁵ Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, D-85748 Garching, Germany
- ³⁶ INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, I-40129 Bologna, Italy
- ³⁷ Dipartimento di Fisica e Scienze della Terra, Università di Ferrara, Via Giuseppe Saragat 1, I-44122 Ferrara, Italy
- ³⁸ Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, I-40126 Bologna, Italy
- ³⁹ Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, PT-1349-018 Lisboa, Portugal
- ⁴⁰ INAF - Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- ⁴¹ Department of Physics and Helsinki Institute of Physics, Gustaf Hållströmin katu 2, 00014 University of Helsinki, Finland
- ⁴² Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
- ⁴³ Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras. 30202, Cartagena, Spain