

Euclid preparation

LV. Exploring the properties of proto-clusters in the Simulated Euclid Wide Survey

Euclid Collaboration: H. Böhringer^{1,2,3,*}, G. Chon^{2,3}, O. Cucciati⁴, H. Dannerbauer⁵, M. Bolzonella⁴, G. De Lucia⁶, A. Cappi^{4,7}, L. Moscardini^{8,4,9}, C. Giocoli^{4,10}, G. Castignani⁴, N. A. Hatch¹¹, S. Andreon¹², E. Bañados¹³, S. Ettori^{4,14}, F. Fontanot^{6,15}, H. Gully¹¹, M. Hirschmann^{16,6}, M. Maturi^{17,18}, S. Mei¹⁹, L. Pozzetti⁴, T. Schlenker¹⁷, M. Spinelli^{7,6,20}, N. Aghanim²¹, B. Altieri²², N. Auricchio⁴, C. Baccigalupi^{15,6,23,24}, M. Baldi^{25,4,9}, S. Bardelli⁴, C. Bodendorf¹, D. Bonino²⁶, E. Branchini^{27,28,12}, M. Brescia^{29,30,31}, J. Brinchmann^{32,33}, S. Camera^{34,35,26}, V. Capobianco²⁶, C. Carbone³⁶, J. Carretero^{37,38}, S. Casas³⁹, F. J. Castander^{40,41}, M. Castellano⁴², S. Cavuoti^{30,31}, A. Cimatti⁴³, C. Colodro-Conde⁴⁴, G. Congedo⁴⁵, C. J. Conselice⁴⁶, L. Conversi^{47,22}, Y. Copin⁴⁸, F. Courbin⁴⁹, H. M. Courtois⁵⁰, A. Da Silva^{51,52}, H. Degaudenzi⁵³, A. M. Di Giorgio⁵⁴, J. Dinis^{51,52}, M. Douspis²¹, F. Dubath⁵³, C. A. J. Duncan⁴⁶, X. Dupac²², S. Dusini⁵⁵, M. Farina⁵⁴, S. Farrens⁵⁶, F. Faustini^{57,42}, P. Fosalba^{41,58}, M. Frailis⁶, E. Franceschi⁴, M. Fumana³⁶, S. Galeotta⁶, B. Gillis⁴⁵, P. Gómez-Alvarez^{59,22}, A. Grazian⁶⁰, F. Grupp^{1,61}, S. V. H. Haugan⁶², W. Holmes⁶³, F. Hormuth⁶⁴, A. Hornstrup^{65,66}, P. Hudelot⁶⁷, K. Jahnke¹³, M. Jhabvala⁶⁸, B. Joachimi⁶⁹, E. Keihänen⁷⁰, S. Kermiche⁷¹, A. Kiessling⁶³, M. Kilbinger⁵⁶, B. Kubik⁴⁸, M. Kümmel⁶¹, M. Kunz⁷², H. Kurki-Suonio^{73,74}, S. Ligorì²⁶, P. B. Lilje⁶², V. Lindholm^{73,74}, I. Lloro⁷⁵, G. Mainetti⁷⁶, D. Maino^{77,36,78}, E. Maiorano⁴, O. Mansutti⁶, O. Marggraf⁷⁹, K. Markovic⁶³, M. Martinelli^{42,80}, N. Martinet⁸¹, F. Marulli^{8,4,9}, R. Massey⁸², S. Maurogordato⁷, E. Medinaceli⁴, Y. Mellier^{83,67}, M. Meneghetti^{4,9}, G. Meylan⁴⁹, M. Moresco^{8,4}, S.-M. Niemi⁸⁴, C. Padilla⁸⁵, S. Paltani⁵³, F. Pasian⁶, K. Pedersen⁸⁶, V. Pettorino⁸⁴, S. Pires⁵⁶, G. Polenta⁵⁷, M. Poncet⁸⁷, L. A. Popa⁸⁸, F. Raison¹, R. Rebolo^{44,89}, A. Renzi^{90,55}, J. Rhodes⁶³, G. Riccio³⁰, E. Romelli⁶, M. Roncarelli⁴, E. Rossetti²⁵, R. Saglia^{61,1}, Z. Sakr^{17,91,92}, A. G. Sánchez¹, D. Sapone⁹³, B. Sartoris^{61,6}, M. Schirmer¹³, P. Schneider⁶⁹, M. Scodeggio³⁶, A. Secroun⁷¹, G. Seidel¹³, S. Serrano^{41,94,40}, C. Sirignano^{90,55}, G. Sirri⁹, L. Stanco⁵⁵, J. Steinwagner¹, P. Tallada-Crespi^{37,38}, A. N. Taylor⁴⁵, I. Tereno^{55,95}, R. Toledo-Moreo⁹⁶, F. Torradeflot^{38,37}, I. Tutusaus⁹¹, T. Vassallo^{61,6}, G. Verdoes Kleijn⁹⁷, A. Veropalumbo^{12,28,98}, Y. Wang⁹⁹, J. Weller^{61,1}, G. Zamorani⁴, E. Zucca⁴, E. Bozzo⁵³, C. Burigana^{100,14}, M. Calabrese^{101,36}, D. Di Ferdinando⁹, J. A. Escartin Vigo¹, F. Finelli^{4,14}, J. Gracia-Carpio¹, S. Matthew⁴⁵, N. Mauri^{43,9}, M. Pöntinen⁷³, C. Porciani⁷⁹, V. Scottéz^{83,102}, M. Tenti⁹, M. Viel^{15,6,24,23,103}, M. Wiesmann⁶², Y. Akrami^{104,105}, V. Allevato³⁰, S. Alvi¹⁰⁶, S. Anselmi^{55,90,107}, M. Archidiacono^{77,78}, F. Atrio-Barandela¹⁰⁸, A. Balaguera-Antolinez^{44,89}, M. Ballardini^{106,4,109}, A. Blanchard⁹¹, L. Blot^{110,107}, S. Borgani^{111,15,6,23}, S. Bruton¹¹², R. Cabanac⁹¹, A. Calabro⁴², F. Caro⁴², C. S. Carvalho⁹⁵, T. Castro^{6,23,15,103}, K. C. Chambers¹¹³, S. Contarini¹, A. R. Cooray¹¹⁴, M. Costanzi^{111,6,15}, B. De Caro³⁶, G. Desprez¹¹⁵, A. Díaz-Sánchez¹¹⁶, S. Di Domizio^{27,28}, H. Dole²¹, S. Escoffier⁷¹, A. G. Ferrari^{43,9}, P. G. Ferreira¹¹⁷, I. Ferrero⁶², A. Fontana⁴², F. Fornari¹⁴, L. Gabarra¹¹⁷, K. Ganga¹⁹, J. García-Bellido¹⁰⁴, T. Gasparetto⁶, V. Gautard¹¹⁸, E. Gaztanaga^{40,41,119}, F. Giacomini⁹, F. Gianotti⁴, A. H. Gonzalez¹²⁰, G. Gozalias^{121,73}, C. M. Gutierrez¹²², A. Hall⁴⁵, W. G. Hartley⁵³, H. Hildebrandt¹²³, J. Hjorth¹²⁴, A. Jimenez Muñoz¹²⁵, J. J. E. Kajava^{126,127}, V. Kansal^{128,129}, D. Karagiannis^{130,20}, C. C. Kirkpatrick⁷⁰, L. Legrand¹³¹, J. Lesgourgues³⁹, T. I. Liaudat¹³², A. Loureiro^{133,134}, J. Macias-Perez¹²⁵, G. Maggio⁶, M. Magliocchetti⁵⁴, C. Mancini³⁶, F. Mannucci¹³⁵, R. Maoli^{136,42}, C. J. A. P. Martins^{137,32}, L. Maurin²¹, R. B. Metcalf^{8,4}, M. Miluzio^{22,138}, P. Monaco^{111,6,23,15}, A. Montoro^{40,41}, A. Mora¹³⁹, C. Moretti^{24,103,6,15,23}, G. Morgante⁴, N. A. Walton¹⁴⁰, L. Patrizii⁹, V. Popa⁸⁸, D. Potter¹⁴¹, I. Risso⁹⁸, P.-F. Rocci²¹, M. Sahlén¹⁴², A. Schneider¹⁴¹, M. Schultheis⁷, M. Sereno^{4,9}, F. Shankar¹⁴³, P. Simon⁷⁹, A. Spurio Mancini^{144,145}, J. Stadel¹⁴¹, S. A. Stanford¹⁴⁶, K. Tanidis¹¹⁷, C. Tao⁷¹, G. Testera²⁸, R. Teyssier¹⁴⁷, S. Toft^{66,148}, S. Tosi^{27,28}, A. Troja^{90,55}, M. Tucci⁵³, C. Valeri⁹, J. Valiviita^{73,74}, D. Vergani⁴, and G. Verza^{149,150}

(Affiliations can be found after the references)

Received 28 July 2024 / Accepted 14 October 2024

* Corresponding author; hxb@mpe.mpg.de

ABSTRACT

Galaxy proto-clusters are receiving increased interest since most of the processes shaping the structure of clusters of galaxies and their galaxy population happen at the early stages of their formation. The Euclid Survey will provide a unique opportunity to discover a large number of proto-clusters over a large fraction of the sky (14 500 deg²). In this paper, we explore the expected observational properties of proto-clusters in the Euclid Wide Survey by means of theoretical models and simulations. We provide an overview of the predicted proto-cluster extent, galaxy density profiles, mass-richness relations, abundance, and sky-filling as a function of redshift. Useful analytical approximations for the functions of these properties are provided. The focus is on the redshift range $z = 1.5$ – 4 . In particular we discuss the density contrast with which proto-clusters can be observed against the background in the galaxy distribution if photometric galaxy redshifts are used as supplied by the ESA *Euclid* mission together with the ground-based photometric surveys. We show that the obtainable detection significance is sufficient to find large numbers of interesting proto-cluster candidates. For quantitative studies, additional spectroscopic follow-up is required to confirm the proto-clusters and establish their richness.

Key words. galaxies: clusters: general – galaxies: high-redshift – large-scale structure of Universe

1. Introduction

Interest in galaxy proto-clusters has recently strongly increased thanks to observational capabilities of new survey instruments, which have revealed how some of the most essential processes shaping the present-day galaxy population in clusters has already happened in the very early stages of cluster formation. The desire to obtain direct observational evidence of these processes at high redshifts has motivated many recent observational studies of proto-clusters (e.g. Overzier 2016; Alberts & Noble 2022).

Galaxy proto-clusters have been found in different ways. They have been found as serendipitous detections in systematic surveys: such as (1) in photometric surveys often conducted to find distant galaxies. One of the first of these discoveries is the proto-cluster in the SSA22 field, which was found as an overdensity of Ly-break galaxies at redshift $z \sim 3$ (Steidel et al. 1998). Other such detections include overdensities of Ly α -emitters (Shimasaku et al. 2003 – Subaru Deep Field, Ouchi et al. 2005 – COSMOS Survey, Higuchi et al. 2019 – SUBARU HSC Survey), i-band drop outs (Toshikawa et al. 2012 – Subaru Deep Field), and multi-band photometric redshifts (Chiang et al. 2014 – COSMOS Survey). Galaxy proto-clusters have also been found (2) in spectroscopic surveys, for example, in the VIMOS Ultra Deep Survey (Cucciati et al. 2014, 2018; Lemaux et al. 2014; Harikane et al. 2019), and (3) as concentrations of sub-millimeter sources in the Planck Survey (Planck Collaboration XXVII 2015; Flores-Cacho et al. 2016; Calvi et al. 2023), by the South Pole Telescope (Vieira et al. 2010; Miller et al. 2018), by the Herschel Space Observatory (Clements et al. 2014; Greenslade et al. 2018), and in other deep survey fields (Daddi et al. 2009; Dannerbauer et al. 2014; Casey et al. 2015; Oteo et al. 2018; Gómez-Guijarro et al. 2019; Wang et al. 2021). Also the detection of proto-clusters as overdensities of passive galaxies has been reported recently (Strazzullo et al. 2015; McConachie et al. 2022; Ito et al. 2023), where the objects of Strazzullo et al. (2015) are not easily classified as either clusters or proto-clusters. However, interesting proto-cluster systems have been discovered around particular objects marking dense regions of the Universe, so-called signposts for proto-clusters, including (4) radio galaxies, which have been used for quite some time to search for dense environments (Le Fevre et al. 1996; Pentericci et al. 2000; Kurk et al. 2000, 2004; Venemans et al. 2002, 2004, 2007; Miley & De Breuck 2008; Kuiper et al. 2011; Hatch et al. 2011a,b; Galametz et al. 2012; Wylezalek et al. 2013; Koyama et al. 2013; Castignani et al. 2014a,b), (5) AGN (active galactic nuclei) playing a similar role as radio galaxies (Djorgovski et al. 2003; Hennawi et al. 2015; García-Vergara et al. 2017, 2019, 2022), (6) and Ly- α blobs (Calvi et al. 2023), and last but not least (7) by absorption in the light of background objects (Francis et al. 1996; Steidel et al. 1998; Hennawi et al. 2015).

While these observed systems span a range of properties (not all of these objects may end up in one galaxy cluster), one looks for a unifying description. The general idea is to call a proto-cluster a structures that is expected to evolve into a galaxy cluster by redshift $z = 0$ (e.g. Steidel et al. 1998; Overzier 2016), whereby theoretical modelling or simulations are used to connect the observations to present-day cluster properties. In Steidel et al. (1998), one of the earliest studies of a proto-cluster, the structure evolution model of a spherical top-hat overdensity was used to relate an observed overdensity to the expectation for a galaxy cluster at the present day. Another approach is to use N -body simulations to trace the evolution of $z = 0$ clusters back to the redshifts of observations and provide, in this way, relations between present-day cluster masses and the properties of their precursors at high redshift. Chiang et al. (2013) provide results from such a study and present correlations between the proto-cluster overdensity and the expected cluster mass at $z = 0$, which has been applied with some success to several proto-cluster observations (e.g. Cucciati et al. 2014). Also, Contini et al. (2016) studied proto-cluster sizes in simulations. However, they used boxes instead of spheres, which makes a comparison to other work more difficult.

The ESA-*Euclid* mission (Euclid Collaboration 2022, 2024a,b,c,d) with its deep near-infrared and high-angular-resolution visual band survey, together with the auxiliary ground-based optical survey data, will provide a unique opportunity to search for proto-clusters over a large region of the sky. This will not only increase the number of known proto-clusters and improve the statistics on their properties, but also yield rare and massive systems, that can only be found in large survey volumes. The main part of the Euclid Survey is the Wide Survey of the sky outside the Galactic band, which has an area of about 14 500 deg² over six years. It will reach estimated limiting AB magnitudes (5σ for point-like sources) of about 26.2 in the visual band, I_E^1 , and 24.5 for the near-infrared bands Y_E , J_E , and H_E (Euclid Collaboration 2022). This paper provides studies of proto-cluster properties and their appearance in the Euclid Wide Survey by means of simulations in order to explore the prospects for the search of proto-clusters in the *Euclid* sky. *Euclid* will enormously increase the number of known proto-clusters at high redshifts and thus provide the base for precise statistical studies on proto-cluster structure, early galaxy evolution in dense environments, and the origin of present day cluster properties. But it will also provide the large survey volume needed to find the most interesting objects, the precursors of the most massive galaxy clusters.

¹ The visual band of the *Euclid* VIS instrument covers a wavelength range from 500 to 900 nm.

One of the major interests in the study of proto-clusters is gaining an understanding of the evolution of galaxies in these dense regions compared to the regions outside of them. The important questions are, when and through which processes was the difference of the more passive present-day cluster members compared to field galaxies established, which epoch saw the bulk of the star formation activity in the cluster galaxies and the most intense enrichment of the intracluster medium with heavy elements, and when and how were the central regions of clusters and the giant central galaxies, such as cD galaxies, formed. An interesting example of the latter question is the study of the Spiderweb proto-cluster (e.g. [Tozzi et al. 2022](#)). The Euclid survey will provide a large statistical sample of proto-clusters, which will not only allow for the investigation of single interesting cases as has been done so far, but enable a study of these questions in the form of population statistics.

In this paper we define a proto-cluster as a matter and galaxy concentration at a higher redshift that is bound to develop into a galaxy cluster by redshift zero with a mass larger than $10^{14} M_{\odot}$ inside r_{200}^2 . We focus mainly on the redshift range $z = 1.5-4$. At higher redshifts, the galaxy density in the Euclid Survey is too sparse to effectively characterise proto-clusters, while at lower redshifts, galaxy clusters are already abundant. In the following, we explore the observational features of such proto-clusters using analytical models and cosmological simulations. We also estimate their abundance as a function of redshift.

The paper is organised in the following way. In Section 2, we describe the formalism of proto-cluster evolution with a top-hat model, while in Section 3 we give information on the cosmological simulations used to explore proto-cluster properties. The following sections are focused on different proto-cluster properties, such as sizes (Sect. 4), galaxy density profiles (Sect. 5), projected contrast in observations (Sects. 6 and 7), and the mass-richness relation (Sect. 8). Section 9 discusses the expected proto-cluster abundances and their sky-filling factors (how much of the sky is covered by proto-clusters in projection). Discussions of the findings are provided in Section 10 and in Section 11 we present our summary and conclusions. Unless stated otherwise, we use a cosmological model with $h_{100} = 0.7 = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and a flat metric, which is referred to as ‘reference model’.

2. Proto-cluster model

As a basic characterisation of proto-clusters, we explore their overdensity evolution in this section. For galaxy clusters and their formation, a simple, general concept that can be expressed with analytic formulas has helped us very much in guiding our thoughts, the so-called Press–Schechter model ([Press & Schechter 1974](#)) and its extensions (e.g. [Bond et al. 1991](#); [Sheth & Tormen 1999](#)). It is in its original form based on the collapse model of a homogeneous overdense sphere and the first-order statistics of density peaks in the large-scale matter distribution. It provides the most essential information to characterise the galaxy cluster population, such as the mass function, their characteristic sizes, and their number density evolution with time. Its precision for the prediction of number counts is usually better than a factor of two for low redshifts and not extremely high masses ($\leq 10^{15} M_{\odot}$) ([Bond et al. 1991](#)). This is not enough for precise cosmological modelling. But this con-

Table 1. Cosmological parameters used in the different models assuming a flat metric.

Model	Ω_m	h_{100}	σ_8
<i>Planck</i>	0.311	0.677	0.816
Millennium	0.25	0.730	0.90
Reference	0.30	0.700	–
REFLEX clusters	0.29	0.700 ^(*)	0.77

Notes. We assume a flat metric. ^(*)The Hubble parameter was pre-set in this study. Its influence on the results is small and discussed in [Böhlinger et al. \(2014\)](#). The REFLEX cluster model is only used in Sects. 9 and 10.

cept has also provided the frame in which more precise analytical models have been devised, which have been calibrated with N -body simulations (e.g. [Jenkins et al. 2001](#); [Evrard et al. 2002](#); [Tinker et al. 2010](#); [Despali et al. 2016](#); [Castro et al. 2021](#)). In recent work, we applied this concept also successfully to those superclusters, which are expected to collapse in the future – calling them “superstes-clusters” ([Chon et al. 2015](#)). These superclusters have the same relation to future galaxy clusters as high-redshift proto-clusters to galaxy clusters of the present day.

Therefore, it is well justified to apply this concept analogously to our proto-cluster project. The adopted model describes the evolution of a homogeneous top-hat spherical overdensity in a Λ CDM universe. The calculation of the evolution is based on Birkhoff’s theorem, where a homogeneous sphere in a homogeneous background universe evolves as a universe would with the local parameters as cosmological parameters. After finding the proper initial conditions for a collapse at redshift zero, we follow the evolution of the overdensity by integrating the Friedmann equations, including a Λ term starting at high redshift to $z = 0$. In the calculation, we determine the overdensity with respect to the mean matter density and the evolution of the radius of the overdense region with respect to r_{200} at $z = 0$.

We define a matter overdensity ratio as the ratio of the mean proto-cluster density, $\bar{\rho}_{pc}(z)$, to the background density, $R_{ov-DM}(z) = \bar{\rho}_{pc}(z)/\rho_m(z)$, where $\rho_m(z)$ is the mean matter density at the given redshift. We have calculated $R_{ov-DM}(z)$ by means of the spherical collapse model for three different sets of cosmological parameters: for the reference model defined in the introduction, the cosmology resulting from the Planck Survey ([Planck Collaboration XIII 2016](#)), and the cosmology used in the Millennium Simulations ([Springel et al. 2005](#)). Table 1 gives these three sets of cosmological parameters. We also list the cosmological parameters inferred from the present-day cluster population in the REFLEX cluster survey ([Böhlinger et al. 2014](#)). The cluster mass function used later in this study ([Böhlinger et al. 2017](#)) is based on these results.

The resulting evolution of $R_{ov-DM}(z)$ is shown in Fig. 1. The curves for the reference model and the *Planck* cosmology can hardly be distinguished in the figure, but the Millennium result is slightly different. The difference is mostly due to the choice of Ω_m and σ_8 , for which an older preference was used in the Millennium Simulations. The difference in the cosmological models does not significantly change the observables discussed in the following. We also indicate in the figure the turn-around redshift, at which the overdensity stops expanding and starts to collapse.

As reference we provide a few numerical values of $R_{ov-DM}(z)$: 3.24, 2.52, 1.91, 1.64 for $z = 1.5, 2, 3, 4$, respectively. The overdensities are small for the time before the turn-around, which occurs at $z \sim 0.8$ in the reference model. For an object

² The radius r_{200} is the radius inside which the mean matter overdensity of the cluster is 200 times the critical density of the Universe at the cluster redshift.

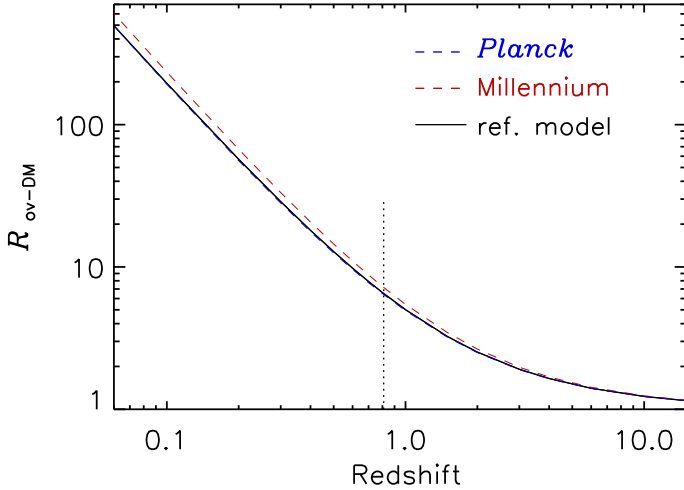


Fig. 1. Evolution of the overdensity ratio, $R_{\text{ov-DM}}(z)$, with redshift for an object that collapses at $z = 0$. The results for three cosmological models are shown: the reference model, Planck cosmology, and Millennium cosmology. The first two models can hardly be distinguished in the plot. The vertical line indicates the redshift of the turn-around point.

that collapses at $z = 0$, the overdensity ratio at turn-around is about 6.6. After the turn-around, the overdensity ratio increases rapidly.

3. Cosmological simulations

Three sets of simulations in the form of lightcones are used for the following study. The MAMBO, GAEA H, and GAEA F lightcones are derived from the Millennium Simulations (Springel et al. 2005; Boylan-Kolchin et al. 2009). The size of the lightcones and the cosmological parameters on which the simulations were based are given in Table 2. Importantly, these lightcones contain information about the merger trees, such that for each galaxy, one can find out if it is included in a larger dark matter halo, a group or cluster of galaxies, at redshift zero. The mass of this final host halo is known, and this parameter is used throughout this paper. These three simulations were the only ones available to us where this information necessary for our work was included. For the two GAEA light cones different semi-analytical models were used to determine the properties of the galaxy population. Some of the following analysis required special information only available for the MAMBO light cone at the time of writing. In this case the results are shown only for this simulation. One of the critical differences in the semi-analytic modelling of the galaxy population that can effect the observables derived from the light cones is the stellar mass function used, as shown by Fu et al. (2024).

The MAMBO simulation based on empirical relations, developed by M. Bolzonella, L. Pozzetti, and G. Girelli, uses the halo and sub-halo positions as well as dark matter masses of one of the 24 lightcones built by Henriques et al. (2015). Under the assumption that each subhalo is hosting one galaxy, the stellar mass was derived from the stellar-to-halo mass relation (Girelli et al. 2020), and the physical and observed properties (e.g. star formation rate, dust attenuation, gas metallicity, morphology, emission lines, broadband rest-frame and observed fluxes) from empirical relations implemented in a modified version of the Empirical Galaxy Generator code (EGG; Schreiber et al. 2017). The predicted galaxy properties also include magnitudes in the H_E and I_E bands. In

Table 2. Parameters for the lightcones and the underlying cosmological simulations.

Data set	Radius (deg)	Sky area (deg ²)	h_{100}	Ω_m	σ_8
MAMBO	1.0	3.14	0.673	0.311	0.816
GAEA H	2.635	21.813	0.73	0.25	0.9
GAEA F	2.635	21.813	0.73	0.25	0.9

Henriques et al. (2015), the Millennium Simulations have been re-scaled to a cosmology consistent with the Planck results (Planck Collaboration XIII 2016; see Table 1).

For the GAEA lightcones, we take advantage of lightcones built from predictions based on the Galaxy Evolution and Assembly (GAEA) theoretical model. This is coupled with dark matter merger trees extracted from the Millennium Simulation (Springel et al. 2005), which makes it possible to track the evolution of each model galaxy both to higher and lower redshifts (in terms of their progenitors and descendants). This approach allowed us to study the later evolution of systems that are identified as proto-cluster regions and characterise the $z = 0$ descendant mass distribution. GAEA follows the evolution of galaxies across different cosmic epochs and environments by means of a coupled system of differential equations, each of them describing a single physical process responsible for the exchange of mass and energy among the different baryonic components. The individual prescriptions can be of empirical, analytical, or theoretical derivation and can involve the definition of free parameters that are usually calibrated against a selected set of observational constraints.

In this paper, we consider two different GAEA realisations based on the model versions published in Hirschmann et al. (2016) and Fontanot et al. (2020). Hereafter, we will refer to these realisations as GAEA-H and GAEA-F, respectively. Both model runs include a detailed treatment for non-instantaneous chemical enrichment (De Lucia et al. 2014) and a prescription for stellar feedback partly based on hydro-simulations. These prescriptions allow us to reproduce the evolution of the galaxy stellar mass function as well as cosmic star formation rate up to $z \lesssim 7$ (Fontanot et al. 2017), as well the evolution of the mass-metallicity relations and their secondary dependencies (De Lucia et al. 2020; Fontanot et al. 2021). GAEA-F also includes an improved treatment of cold gas accretion onto supermassive black holes, an explicit treatment for AGN-driven winds (Fontanot et al. 2020), and an updated tracing of the angular momentum exchanges between different galactic components, which we use to model galaxy structural properties (Xie et al. 2020). While retaining all successes of the previous model, this update also reproduces the properties of the AGN populations up to $z \sim 4$.

For each model galaxy, GAEA predicts the expected broadband photometry in the H -band and in the *Euclid* visual band, in addition to a number of physical properties and other photometric bands. These model outputs have been used to construct two independent lightcones (each for each model realisation) using the algorithm described in Zoldan et al. (2017). Each lightcone covers an aperture of $5:27$ diameter and includes all model galaxies from $z = 0$ to $z = 4$ down to $H_E = 25$.

In the simulations, galaxies keep their identity through time and can be traced through the merger trees. Therefore, galaxies that are found to belong to a galaxy cluster at redshift zero with

a mass $M_{200} \geq 10^{14} M_{\odot}$ inside r_{200} can be identified and labelled as members of a proto-cluster at higher redshift in the lightcone. The ensemble of the member galaxies defines the proto-cluster.

For our studies, we selected galaxies with the following magnitude limits, $m_{H_E} \leq 24.25$ for the near-infrared H -band and $m_{I_E} \leq 25.10$ for the visual band of *Euclid*³. The lightcone databases provide observed and randomly perturbed magnitudes according to the expected measurement errors. Here, we use the unperturbed magnitudes. The limits correspond to an expected $S/N=5$ *Euclid* will reach in the Wide Survey for extended sources. For each galaxy, redshifts calculated from the simulations with and without peculiar motions are available. We use the values without peculiar motion. Also, the photometric redshift was derived for each galaxy with the SED fitting code Phosphoros developed in the Euclid Collaboration (Paltani et al., in prep.), taking into account the photometric noise expected in *Euclid* bands complemented by the ground-based ones that will be available at the time of the Data Release 3 southern hemisphere⁴ (Euclid Collaboration 2022). The lightcone databases contain the entire probability distribution of the derived photometric redshift. Here, we use only the median values.

4. Extent of proto-clusters

In this section, we present estimates of the proto-cluster radius, r_{pc} . We first derive a theoretical estimate for the radius evolution of a spherical overdensity in Sect. 4.1 and compare the results to simulations in Sect. 4.2.

4.1. Theoretical estimate

For r_{pc} , we took the radius of the overdensity, which evolves into a galaxy cluster with radius r_{200} in the spherical collapse model described in Sect. 2. The proto-cluster radius, r_{pc} in comoving units, r_{com} , can then be determined from the overdensity ratio through the relation:

$$r_{com} = r_{200} \left(\frac{\Omega_m R_{ov-DM}}{200} \right)^{-1/3}, \quad (1)$$

and the physical radius of the proto-cluster is given by $r_{phys} = r_{com}/(1+z)$.

Figure 2 shows the results of the calculations for the three cosmological models listed in Table 1. The results for the reference and *Planck* cosmology are practically identical, while those for the Millennium cosmology are different by a few per cent. Chiang et al. (2017) published similar calculations based on simulations, which are in good agreement with our results. Similarly, Muldrew et al. (2015) used simulations to assess the evolution of the proto-cluster radius, defined as enclosing 90% of the stellar mass ending up in the $z=0$ cluster and found similar results. In their Fig. 2, we see that the physical radius shows a similar function of time with a value ~ 2.5 times higher at $z=1$ than at $z=0$ and a factor of ~ 1.2 higher at $z=5$.

For further practical work, we derived numerical fits to the results for r_{com} . The following approximation, valid for the red-

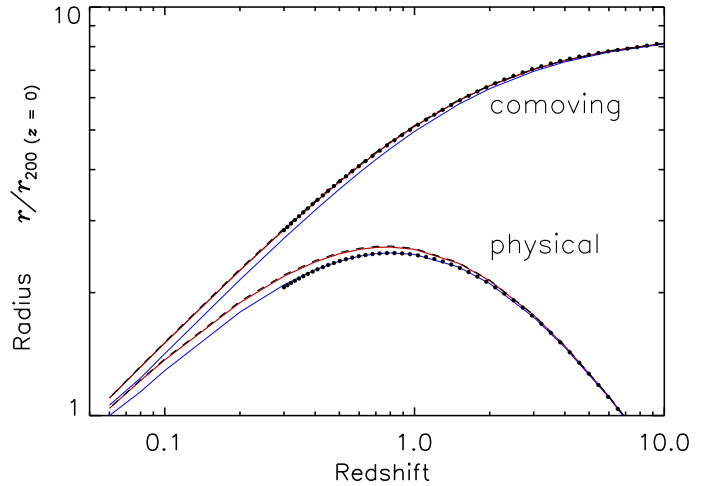


Fig. 2. Evolution of the proto-cluster radius in comoving and physical units scaled to r_{200} at $z=0$. Results for three cosmological models are shown by red lines (reference model), blue lines (Millennium cosmology) and dashed black lines (Planck cosmology). The over-plotted black dots show the parameterised approximation of r_{com} (see Eq. (2)) for the reference cosmology and r_{phys} for the Millennium cosmology.

Table 3. Parameters describing the evolution of the proto-cluster radius with redshift (see Eq. (2)).

Model	A	B	C	D
Reference	19.50	-22.75	9.05	-0.71
<i>Planck</i>	19.54	-22.83	9.08	-0.68
Millennium	19.38	-22.43	8.88	-0.88

Notes. The approximations are only valid in the redshift range $z=0.5-8$.

shift range $z=0.5-8$, provides an accuracy better than 1%, with parameters listed in Table 3,

$$r_{com}/r_{200} = A z^{0.7} + B z^{0.95} + C z^{1.1} + D. \quad (2)$$

We use this relation for the reference cosmology in the subsequent work in this paper.

4.2. Comparison to simulations

We can compare the theoretical predictions for the proto-cluster radius with the distribution of the cluster member galaxies in the MAMBO and GAFA lightcones. Here and in the following, we use only proto-clusters that are fully contained in the field of view of the lightcone and in addition reject a small number of objects that are artefacts originating from common problems with the lightcone construction.

Figure 3 shows two proto-clusters from the MAMBO simulation, where galaxies inside and outside the proto-cluster radius (in three dimensions) are marked as members and non-members. As the centre of the proto-clusters, we chose the barycentre. As proto-cluster members, we took those galaxies that are members of the descendent cluster at redshift zero inside r_{200} . All galaxies in a redshift slice with a width of five times the proto-cluster radius are shown in projection on the sky. We note that the majority of the proto-cluster members are located inside the estimated proto-cluster radius, and a few non-members, marked in blue in the figure, are found inside. These non-members are found close to the cluster boundary and are not bound into the cluster during

³ We note that these magnitude limits deviate slightly from the limits given for the survey in Sect. 1. The constantly evolving work on the survey definition and in the science working group cannot always be perfectly synchronised in the Euclid preparation phase.

⁴ The photometric data used for the determination of the photometric redshifts include the *Euclid* bands: I_E , Y_E , J_E , H_E and the LSST bands: u , g , r , i , z .

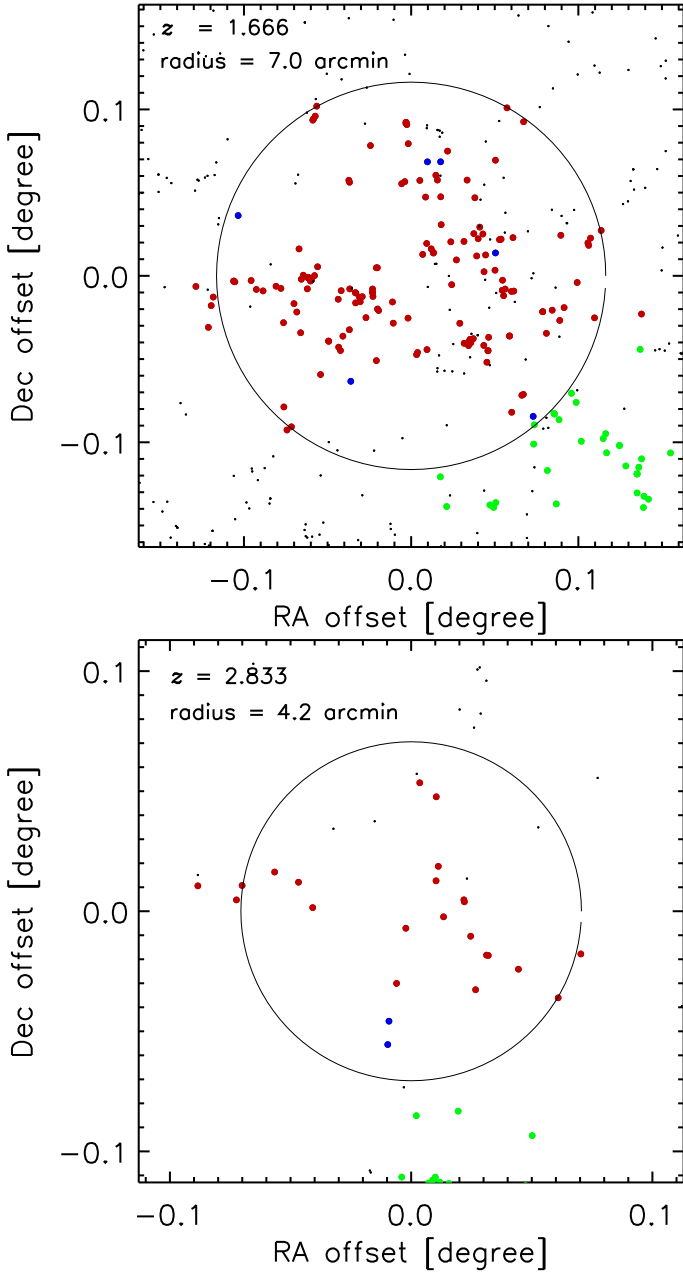


Fig. 3. Examples of two proto-clusters at redshifts 1.666 and 2.8335 with 148 and 22 members, respectively. The proto-cluster members are shown as red-filled circles, while non-members inside the proto-cluster radius are shown as blue circles. Other non-members outside the proto-cluster radius in three dimensions are shown as small black dots, and members of other proto-clusters are shown as full green circles. The large black circle indicates the estimated proto-cluster radius, whose size is indicated in the top left corner of the plot.

the following collapse of the system. Some of the non-members are seen inside the circle in projection in the plots as black dots, but they are outside the proto-cluster spheres. Members of neighbouring proto-clusters in the same redshift interval are shown as green symbols.

Figure 4 shows the three-dimensional radial distribution of the galaxy number and number density of the members of MAMBO, GAEA-H, and GAEA-F proto-clusters compared to the radial distribution of non-members. The differential distribution in spherical shells is shown. For the proto-cluster centres,

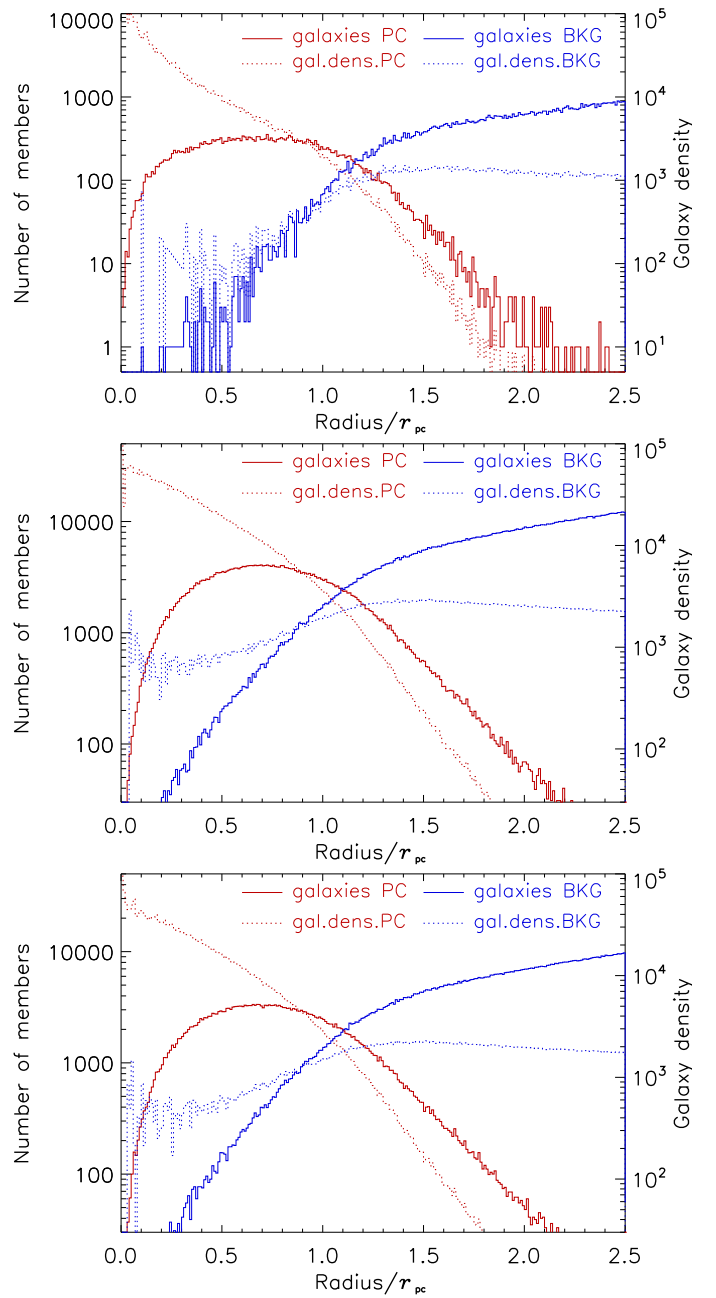


Fig. 4. Three-dimensional radial distribution of the member galaxies in all MAMBO (top), GAEA-H (middle), and GAEA-F (bottom) proto-clusters ($z = 1.5-4$). The radius is in units of the estimated proto-cluster radius, r_{pc} . The red lines show the numbers and densities of the proto-cluster members, and the blue lines those of the non-member galaxies. The solid lines show the galaxy number in the shells (left Y-axis), while the dotted lines show the galaxy density in the shells (right Y-axis).

we used the barycentre, the mean of the mass distribution of all proto-cluster member galaxies. The distance of the galaxies to the centre in all proto-clusters was scaled to the estimated proto-cluster radius, $r_{pc} = r_{phys} r_{200}$. The radius, r_{phys} , has been defined in the sentence connected to Eq. (1). For the positions, the true locations of the galaxies were used without redshift space distortions due to galaxy peculiar motions.

The distributions of the galaxies in all three simulation data sets look very similar. The contamination of non-members inside the proto-cluster radius in the GAEA lightcones is higher than

Table 4. Completeness and contamination of the members inside the estimated proto-cluster radius, r_{pc} and $1.2 r_{\text{pc}}$, for different redshift shells for the MAMBO and GAEA samples

Redshift	Compleat. $r \leq r_{\text{pc}}$	Contam. $r \leq r_{\text{pc}}$	Compleat. $r \leq 1.2 r_{\text{pc}}$	Contam. $r \leq 1.2 r_{\text{pc}}$
GAEA H				
1.5–4	0.74	0.13	0.88	0.21
1.5–2	0.70	0.17	0.84	0.25
2–3	0.76	0.12	0.89	0.20
3–4	0.80	0.08	0.91	0.16
GAEA F				
1.5–4	0.75	0.12	0.88	0.21
1.5–2	0.70	0.16	0.84	0.25
2–3	0.76	0.11	0.90	0.20
3–4	0.81	0.07	0.90	0.15
MAMBO				
1.5–4	0.78	0.043	0.91	0.11
1.5–2	0.76	0.053	0.90	0.12
2–3	0.79	0.034	0.91	0.11
3–4	0.81	0.038	0.92	0.11

that in the MAMBO lightcone. About 80% of the members are located inside r_{pc} with a contamination of $\sim 5\%$ (MAMBO), $\sim 10\%$ (GAEA) non-members in three dimensions. About 90% of the members are located inside $1.2 r_{\text{pc}}$ with a contamination of $\sim 10\%$ (MAMBO) and $\sim 20\%$ (GAEA). At the moment we have no explanation for this difference of the contamination in the two light cones. Apart from this effect we see no significant difference between the different simulation samples. The theoretically estimated proto-cluster radius thus provides a good orientation for the expected size of proto-clusters. Results for splitting the proto-cluster sample up into three redshift shells are given in Table 4, where the completeness is defined as the fraction of member-galaxies contained inside the aperture radius compared to the total number of member galaxies. We note little variation with redshift.

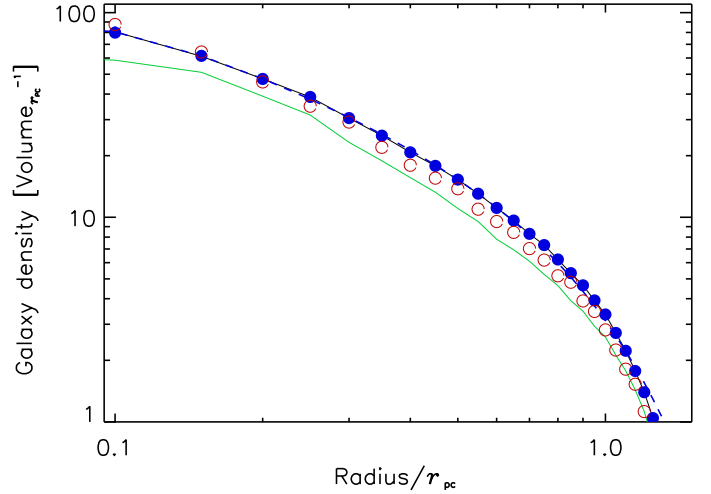
5. Radial galaxy density profiles

We used the GAEA and MAMBO simulations to study the typical radial density profiles of proto-clusters. Here and throughout the paper, we use the barycentre of the galaxy population as the proto-cluster centre. In Fig. 5 we show the mean three-dimensional proto-cluster profile for the GAEA-F and MAMBO samples for the redshift range $z = 1.5-4$. In addition we show with data for GAEA-F in the range $z = 1.5-2$, that there is no substantial change with redshift. The profiles were generated involving 6533 and 786 proto-clusters from the two simulations, respectively. Since we used radii scaled to the proto-cluster radius, r_{pc} , we also determined the mean densities as scaled parameters in units of $4/3\pi r_{\text{pc}}^3$. The mean profile for GAEA-H has the same shape but about 10% higher normalisation since this sample has relatively more proto-cluster members.

We fitted the resulting profile with a model including a power-law profile with a core and an additional inner and outer slope:

$$n(r) = A \left(\frac{r}{r_c}\right)^{-\alpha} \left[1 + \left(\frac{r}{r_c}\right)^2\right]^{-\beta + \frac{\alpha}{2}} \left[1 + \left(\frac{r}{r_s}\right)^2\right]^{-\gamma + \beta + \frac{\alpha}{2}}. \quad (3)$$

The resulting fit parameters for the three-dimensional case are shown in Table 5. We also determined the projected two-dimensional profiles, which look similar with a flatter slope. The

**Fig. 5.** Mean three-dimensional radial density profile of proto-clusters. The black solid line and blue points show the mean profile for all proto-clusters in the GAEA F lightcone in the redshift range $z = 1.5-4$, while the green curve shows that for proto-clusters with $z = 1.5-2$. The blue dashed curve shows the model fit. The red open symbols show the density profile for the clusters in the MAMBO lightcone ($z = 1.5-4$) for comparison. The galaxy density is scaled to the volume of the proto-cluster, $\text{volume}_{r_{\text{pc}}} = (4/3)\pi r_{\text{pc}}^3$.**Table 5.** Fit parameters for the proto-cluster profiles in three and two dimensions for the formula in Eq. (3).

Sample	A	r_c	α	β	r_s	γ
3 dim.						
GAEA H	117.7	0.1413	0.2163	0.635	11.35	115.9
GAEA F	187.9	0.0625	-0.989	0.530	15.74	286.6
MAMBO	55.32	0.2371	0.6368	0.790	16.71	198.3
2 dim.						
GAEA H	34.79	1.004	0.2121	2.98	242.2	1050
GAEA F	59.09	0.0968	-0.110	0.254	15.61	364.0
MAMBO	127.4	0.0209	-2.940	0.279	9.475	128.8

Notes. The radii, r_c and r_s , are in units of r_{pc} .

profiles were fitted with the same relation, and the results are also shown in Table 5. In this case, r_c and r_s are projected radii. Figure 5 shows that the fit provides reasonable approximations. Actually, a fit with a function including only the first four parameters, with the first two terms on the right hand side of Eq. (3), provides a good approximation up to $r \sim r_{\text{pc}}$, but leaves a small shallow tail beyond. To remove this particular feature with an additional outer slope requires large values for the core radius, r_s , and the slope parameter.

One interesting result from the average projected proto-cluster profile to keep in mind is that about half of the member galaxies reside inside $0.5 r_{\text{pc}}$. That implies that the density inside an aperture of $0.5 r_{\text{pc}}$ is about 3 times higher than the density in the annulus at $r = 0.5$ to $1 r_{\text{pc}}$.

Inspection of individual profiles shows that there is a large variety of profile shapes. Figure 3 and in the Appendix Fig. A.3 show a selection of proto-clusters with some emphasis on systems with substructure. To devise a cluster detection method that works with assumptions on the proto-cluster shapes, as, for example, a matched filter algorithm, one needs an overview of the variation of the proto-cluster structures. Any treatment that

Table 6. Definition of the five different categories of proto-cluster density profiles.

Category	Condition	N_{pc}	Percent
1	D1 > D2 > D3	542	69
2	D1 > D3 > D2	33	4.2
3	D2 > D1 > D3	67	8.5
4	D2 > D3 > D1	127	16.1
5	D3 > D1 and D2	17	2.2

Notes. D1, D2, and D3 are the mean galaxy densities in the three radial intervals, 0 to 1/3, 1/3 to 2/3, and 2/3 to $1 r_{\text{pc}}$, respectively. The last two columns give the number of proto-clusters and the percentage in each category.

regards azimuthally symmetric shapes in first order would rely on knowledge of profiles.

To get such an overview, we dissected the three-dimensional profiles out to r_{pc} into three equal radial intervals and classified the proto-cluster profiles into five categories according to the density ratios of the different regions as listed in Table 6. Category 1 and 2 have a central peak, 3 and 4 have the maximum in the middle, and 5 has the maximum near r_{pc} . Table 6 also shows the number of proto-clusters that fall into each category.

To display the variation of the profiles, we have determined the mean profile for each category with a resolution of 8 bins out to $1.6 r_{\text{pc}}$. We note that this provides a higher radial resolution of 5 bins inside r_{pc} than the three radial bins used for the categorisation. This allows us to show the profiles in more detail. Not to let those proto-clusters with the largest number of galaxy members completely dominate the results, we use a weighting with a factor of $n_{\text{gal}}^{-1/2}$ per proto-cluster in averaging the profiles, where n_{gal} is the total number of members in the simulations. Due to the scaling with proto-cluster radius and this weighing scheme, the resulting densities lose the normal physical units, and we show relative values. These mean profiles for categories 1–5 are shown in Figs. A.1 and A.2 in the Appendix. We note first of all that 73.2% of the proto-clusters have a high central density inside $0.2 r_{\text{pc}}$ (much higher than in the other radial bins). 4.2% of these proto-clusters have a higher density in the third than in the second bin. But as shown in Fig. A.1 they have mostly very compact cores with few galaxies outside. The 24.6% of category 3 and 4 have the highest density in the middle radial region. Two examples of such clusters are shown in the upper two panels of Fig. A.3.

Only 17 proto-clusters have the highest density in the region from $2/3 r_{\text{pc}}$ to $1 r_{\text{pc}}$. An example is shown in the bottom panel of Fig. A.3. Among these cases with high density in the outer annuli we find binary and multiple systems, which will nevertheless collapse into a single cluster by $z = 0$. Thus, the majority of the proto-clusters will show a significantly higher contrast of the central region compared to the overall system. It also implies that the central region will collapse earlier than redshift zero for most proto-clusters to form a smaller galaxy group or cluster first.

6. Density contrast in the projected galaxy distribution

6.1. Density contrast with respect to the global sky background

To explore the efficiency with which proto-clusters can be detected in the Euclid Wide Survey, we study the projected galaxy density contrast of proto-clusters against the galaxy back-

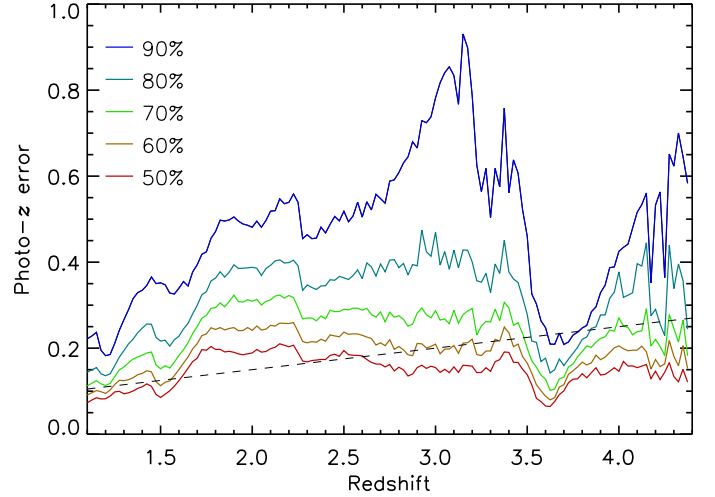


Fig. 6. Photo- z error as a function of redshift for different completeness limits. The parameter photo- z error describes the half-width of the redshift window necessary to contain 50%, 60%, 70%, 80%, and 90% of all galaxies at a certain redshift. The dashed line shows the official requirement for the photometric accuracy.

ground in this section. Here and in the following we designate all galaxies outside the proto-clusters, that is either in front or in the back, as background galaxies. We make use of the photometric redshifts that have been modelled in the simulations. To find the majority of the members of a proto-cluster we have to consider a selection window for the redshifts wide enough to cover the uncertainties of the photometric redshifts. Using the MAMBO lightcone, we illustrate in Fig. 6 the median photometric errors corresponding to including 50–90% of all galaxies at a given redshift, which we define as ‘completeness’. These photometric errors constitute the redshift windows used to detect PC. The half window size is shown as photo- z error parameter as a function of redshift in the Figure. We will designate the redshift interval for a 50% completeness limit as $\Delta z_{(50\%)}$. These calculations have been performed for all galaxies in the lightcone, including non-members, for better statistics in 132 redshift bins. In each bin, the galaxies were sorted by Δz , the deviation of the photometric redshift from the true one. The maximum Δz of the 50% smallest values yields then, for example, $\Delta z_{(50\%)}$. This completeness limit is identical to the median. We used this value in most of the following examples because it provides a detection efficiency close to the maximum, as shown in Sects. 6.2 and 7. Also shown is the official *Euclid* requirement for the redshift accuracy, $\Delta z = 0.05(1+z)$ (e.g. Sartoris et al. 2016). This requirement is close to the median curve. At redshifts between $z = 1.5$ and 2.2 , it is worse than the requirement. With increasing redshift, it gets better and falls below the requirement at $z = 3-4$. At these higher redshifts, the photometric bands bracket the Ly-break better.

Galaxy counts and densities in proto-clusters and background in the MAMBO lightcone (limited by the redshift range corresponding to given redshift uncertainties) were calculated for different aperture sizes (with respect to r_{pc})⁵. In practice, we counted all known proto-cluster members inside the aperture radius and all other galaxies as background in the cylinders defined by the redshift uncertainty, $\Delta z_{(X\%)}$, and the aperture

⁵ We present here only the results from the MAMBO lightcone due to some problems in the cross-identification of individual proto-cluster members in different simulation result tables at the time of writing. This problem is being solved.

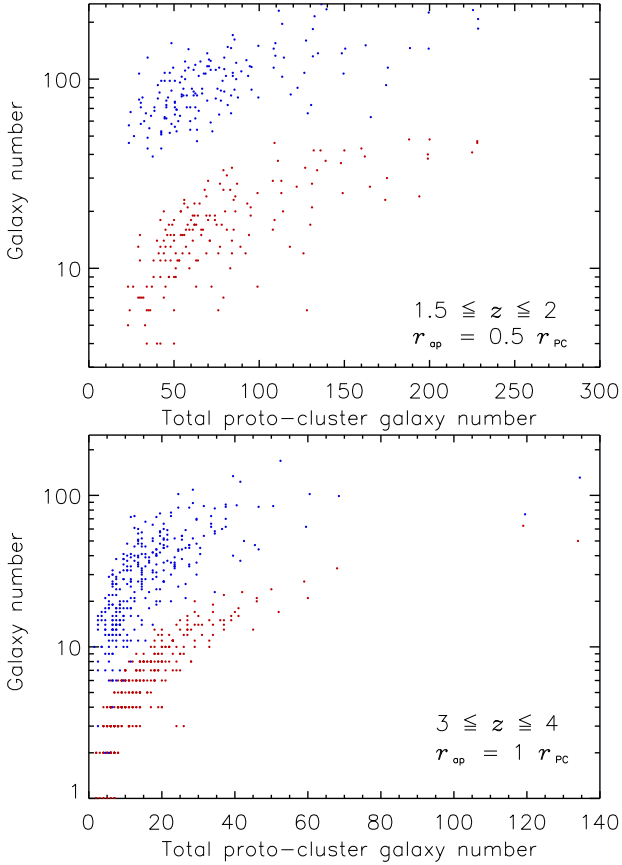


Fig. 7. Galaxy counts in proto-clusters (red) in comparison to the background (blue) as a function of the total galaxy number in the proto-clusters. The top row was taken in the redshift range $z = 1.5-2$ with an aperture of $0.5 r_{\text{pc}}$ and the bottom row in the range $z = 3-4$ inside an aperture of r_{pc} . For the galaxy selection a redshift interval corresponding to $\Delta_z (50\%)$ was used.

area. This was done for different redshift regions and different photometric uncertainty limits. Figure 7 shows two examples of the galaxy counts in proto-clusters and background for apertures, $r_{\text{ap}} = 0.5 r_{\text{pc}}$ for $z = 1.5-2$ and $r_{\text{ap}} = r_{\text{pc}}$ for $z = 3-4$ using a redshift window defined by $\Delta_z (50\%)$. We note that the background galaxies always outnumber the proto-cluster galaxies. This is due to the fact that background galaxies are sampled over a much larger line of sight than proto-cluster members due to the limited photometric redshift accuracy, as will be explained below. We show the data as a function of the total galaxy number of the proto-cluster, which is a good proxy for the proto-cluster mass (Rykoff et al. 2014). Of course, the number of recovered proto-cluster galaxies increases with the total galaxy number, but also, the background increases somewhat due to the increase of r_{pc} . Statistically, the number of detected galaxies with large aperture radius ($r \sim 2.5 r_{\text{pc}}$) scatters around 50% of the total number by construction for a $\Delta_z (50\%)$ limit.

Results for the densities of the detected galaxies are shown in Fig. 8. Again, the projected densities of the background are always larger than the proto-cluster densities. As expected the galaxy density in the background is constant, while the average projected galaxy density increases with proto-cluster size and richness, as the proto-cluster volume increases. The latter increase is mainly shown by an increase in the mean density and its lower limit. Figure 9 summarises the results for the density statistics as a function of aperture radius and redshift for a 50%

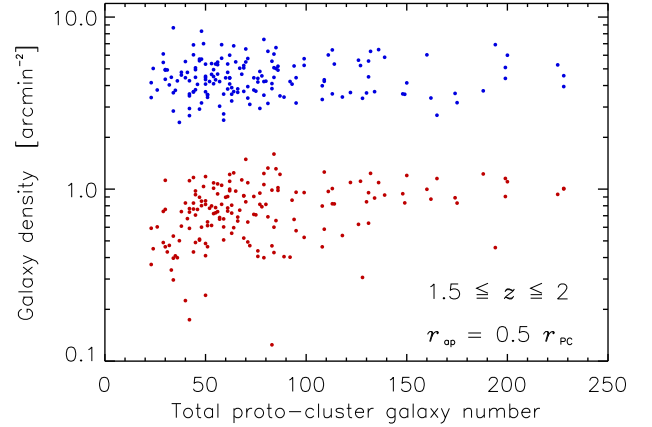


Fig. 8. Galaxy densities in proto-clusters (red) in comparison to the background (blue) in the redshift range $z = 1.5-2$ with an aperture of $0.5 r_{\text{pc}}$. For the galaxy selection a redshift interval corresponding to $\Delta_z (50\%)$ was used.

completeness and *Euclid* magnitude limit. Here we note that the background densities appear constant with changing aperture radius as expected. The densities for the proto-clusters, however, decrease with increasing aperture radius due to the decreasing density profiles.

The fact that the projected background density is often very much larger than the proto-cluster density is a challenge for the reliable detection of the proto-clusters. Thus, before proceeding further, we study the reason for this situation in more detail. We have shown in Sect. 2 that the proto-cluster overdensities are not very large since we capture the proto-clusters before turnaround. This overdensity has to be compared with the line-of-sight ratio across the proto-cluster and the redshift range defined by the photometric redshift accuracy. In projection, we sample all the galaxies in the line-of-sight of the proto-cluster, which have a redshift inside the uncertainty limits of the photometric redshifts. Since the line of sight distance interval corresponding to the redshift uncertainties is much larger than the diameter of the proto-clusters, more background galaxies are sampled than proto-cluster members by the photometric redshift selection, in spite of the moderate galaxy overdensity in proto-clusters. To illustrate this, we show in Fig. 10 the ratio between the distance interval corresponding to the photometric redshift uncertainty and the proto-cluster diameter as a function of redshift. Here, we have used a redshift interval of $\Delta_z (50\%)$; with a smaller completeness, the volume ratio would be smaller, but we would also sample fewer member galaxies. We clearly see that the volume from which the background galaxies are sampled is much larger than the proto-cluster volume (which has been simply approximated here by a cylinder). This large volume ratio is the consequence of using only broadband photometry for the redshift estimates.

6.2. Significance of the density contrast

An additional problem complicates the detection of proto-clusters. While it is sometimes assumed naively that one detects the proto-clusters in a distinct and smooth background field, which is characterised by Poissonian density fluctuations, we are facing the situation that we have to detect the proto-clusters against a background characterised by the presence of cosmic large-scale structure. Since the aperture sizes for detecting proto-clusters sample the background at relatively small scales, where the cosmic large-scale structure is well in the non-linear regime,

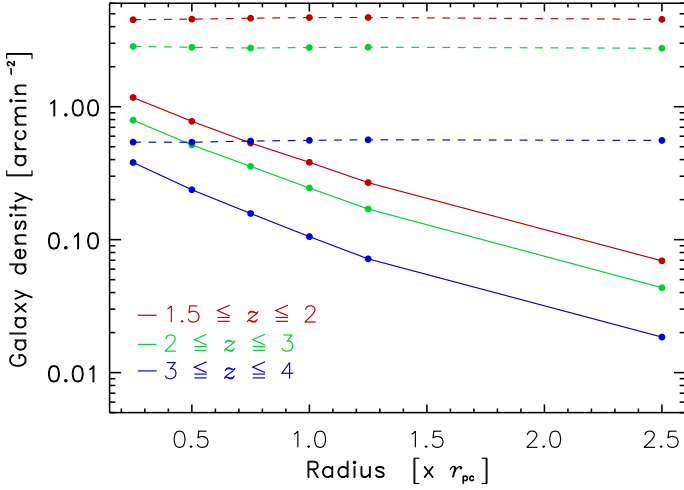


Fig. 9. Mean projected galaxy densities in proto-clusters (solid lines) as a function of aperture radius in three redshift intervals. The dashed curves give the galaxy densities in the background. The results were derived for $\Delta_z(50\%)$.

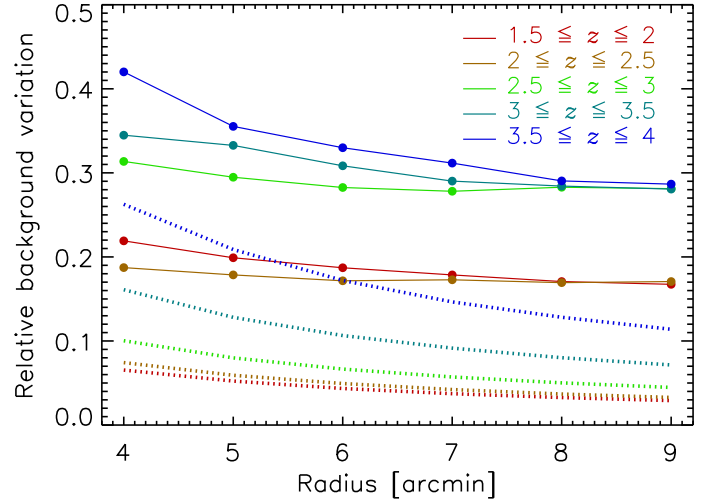


Fig. 11. Relative rms deviations of the projected background galaxy density fluctuations (solid lines) as a function of aperture radius for five different redshift intervals. The dashed lines show the estimated Poisson errors for comparison.

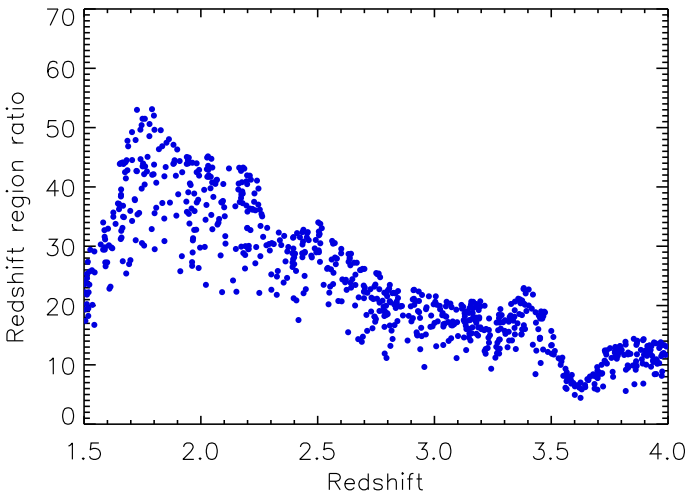


Fig. 10. Ratio between the line-of-sight interval given by the photometric redshift uncertainty and the line of sight covered by proto-clusters, $r(\Delta_z(50\%))/r_{pc}$. The ratios for all MAMBO proto-clusters as a function of redshift are shown.

we have to cope with background fluctuations larger than the Poisson noise for the relevant galaxy counts. We illustrate this by means of the MAMBO simulation below.

The rms of the background density fluctuations was determined in different aperture radii in five redshift ranges ($z = 1.5-2$, $2-2.5$, $2.5-3$, $3-3.5$, and $3.5-4$). The background densities were evaluated at the positions of the proto-clusters, while the proto-cluster galaxies were excluded from the background density calculation. In principle, one could have also used random positions, but in our approach, we include the small effect that a tiny part of the line of sight is occupied by the proto-cluster and not by the background. This time the aperture radius is kept fixed for a chosen aperture value and not scaled with the proto-cluster radius. For the line of sight integration we take the redshift interval $\Delta_z(50\%)$. We show the results in Fig. 11 for six aperture radii (4, 5, 6, 7, 8, and 9 arcmin) and compare them with Poisson errors. While the Poisson errors decrease steadily with the aperture radius, as expected, we note hardly any decrease

in the variance, except for the first aperture radius bin, where shot noise still plays an important role. The variance decreases with aperture radius not because of the improving count statistics but because of the decreasing variance of the large-scale structure with scale. This decrease is comparatively slow and hardly noticeable in the relevant scale range. For this reason, the significance of detection above the background cannot easily be improved by increasing the aperture size, as it would be for Poisson errors.

To characterise the significance of detection of the proto-clusters we take the ratio of the detected counts to the rms of the background fluctuations. We do not include the shot noise of the proto-cluster counts, which would be important in the calculation of the error of the detected signal. We look here just at the detection significance. We determine the detection significance for every proto-cluster by using the actually detectable counts in the simulations, and for the background, we take the rms of the background counts determined from all proto-cluster positions in the same redshift bin and for the same aperture radius. We note that the fluctuations on the relevant scales are actually mildly non-linear and therefore the use of the RMS for the background noise is only a reasonable approximation. This will be accounted for more precisely in the practical application with the calibration of the proto-cluster detection techniques in simulations. Examples of the results are shown in Fig. 12 for the redshift range $z = 1.5-2$ for aperture radii of $r_{ap} = 0.5$ and $1.0 r_{pc}$. We show the detection significance as a function of the total galaxy number belonging to the proto-clusters in the simulation.

The values for the significance are small. There is a correlation of the significance with the richness of the proto-cluster. This correlation is more pronounced for the better number statistics with the larger aperture. For the smaller aperture the behaviour for the less rich proto-clusters is dominated by Poisson noise, but a correlation becomes apparent at richer proto-clusters: the mean and the lower limit of the significance clearly increases.

The upper panel of Fig. 13 summarises the significance studied for a $\Delta_z(50\%)$ limit. In the plot, we show both the mean and maximum values for the significance of each parameter selection. The maximum values for the smallest aperture radius are

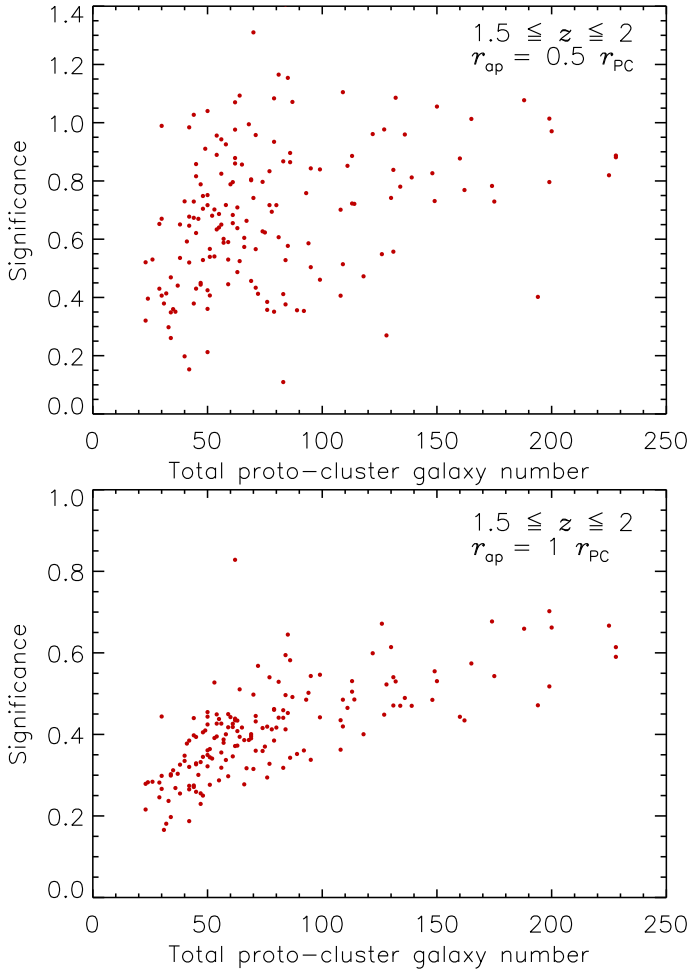


Fig. 12. Detection significance of proto-clusters from the MAMBO simulation in the redshift interval $z = 1.5\text{--}2$. The panels show results for aperture radii of 0.5 (top) and $1 r_{\text{pc}}$ (bottom). For the galaxy selection a redshift interval corresponding to $\Delta_z (50\%)$ was used.

relatively high due to the large Poisson noise for small counts. We note that for the highest redshift bin, the maximum values of the significance is always highest for the same reason, but for the mean values, the highest redshift bin is not more significant than the lowest one.

We also explored if we can increase the significance by using a higher or lower completeness limit so that we can sample more galaxies for each proto-cluster or deal with a smaller redshift range for the background. The lower panel of Fig. 13 shows how the significance changes with the completeness limit for two apertures (0.5 and $1 r_{\text{pc}}$) and three redshift intervals. We note that the 50% or 60% completeness limits provide the maximum in all cases. Otherwise, the curves are relatively flat. Only for the highest redshift bin there is a notable decrease of the significance with increasing completeness because it is closer to the detection limit for the galaxies. The main conclusion is, however, that there is not much room for improvement in the detection efficiency with a different choice of the completeness limit.

Overall, a strategy that uses $\Delta_z (50\%)$ and apertures smaller than $1 r_{\text{pc}}$ (for example $0.5 r_{\text{pc}}$) provides a close-to-optimal solution. While at decreasing radii the signal-to-noise gets better, the statistics gets worse, and therefore a radius of around $0.5 r_{\text{pc}}$ gives a good compromise. Thus for most of the following analysis we use this selection.

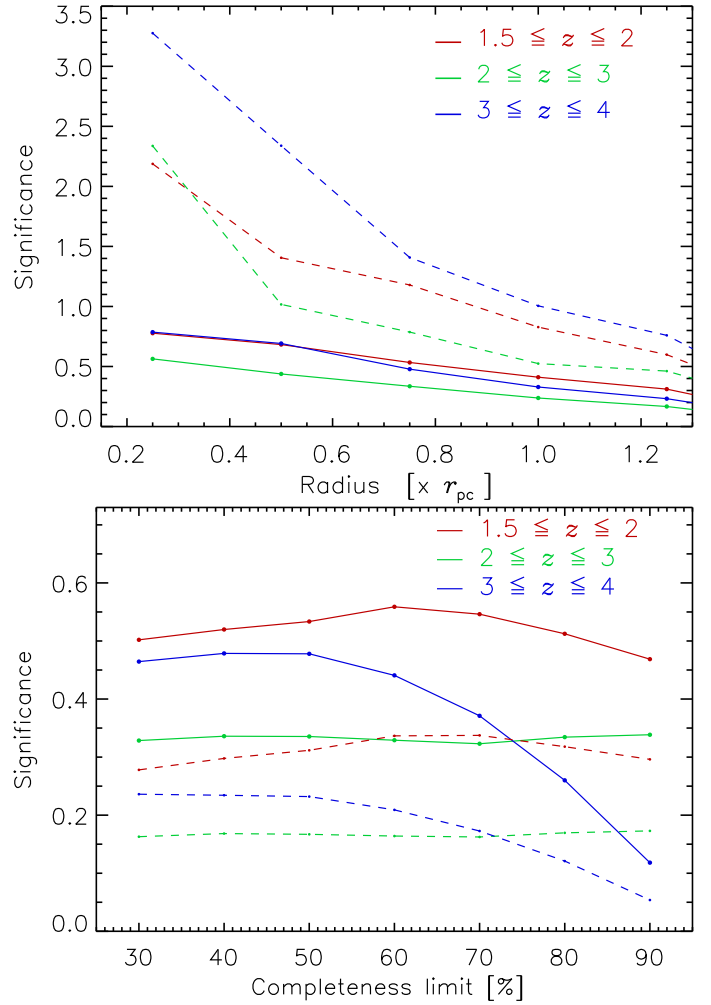


Fig. 13. Detection significance of proto-clusters as a function of aperture radius (top) and as a function of completeness limit (bottom) for three different redshift intervals. Top: The solid curves show the mean significance, while the dashed curves show the maxima. A redshift interval of $\Delta_z (50\%)$ was used. Bottom: The completeness limit, $\Delta_z (X\%)$, covers the range $30\text{--}90\%$. The solid lines show the results for an aperture radius of $0.5 r_{\text{pc}}$ and dashed lines for $1 r_{\text{pc}}$.

An improvement can, of course, be expected from using a probability-based detection algorithm, for example, a matched filter technique. The present study provides a useful guideline for such a method since it shows which proto-cluster region contributes most to the detection signal.

7. Detection significance with local background assessment

In this section we explore whether the detection significance can be improved with a local background assessment. In the previous section, we used the global background variance for the significance calculation. Here, we test the behaviour of the background if the background is taken for each proto-cluster from an annulus around it and its variance determined from the data. Due to the spatial correlation of the galaxy density in the large-scale structure, we could expect that there would be a correlation between the background galaxy density inside the proto-cluster aperture and that of the surrounding annulus. This correlation is shown for the simulated MAMBO proto-clusters in Fig. 14. Since there

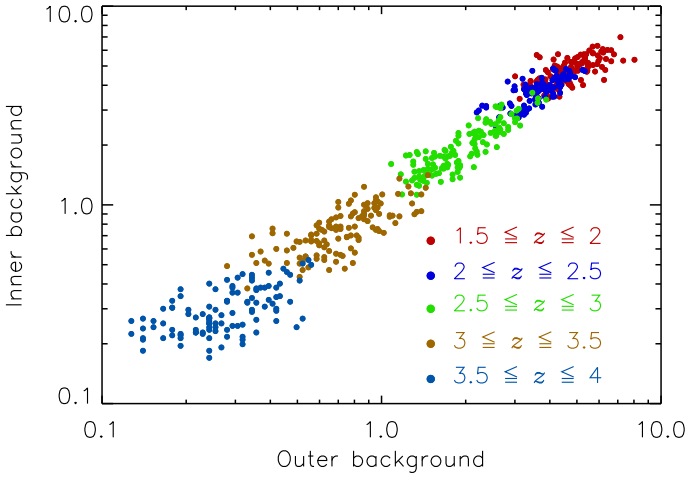


Fig. 14. Correlation of the galaxy background density for proto-clusters in an aperture radius of 5 arcmin and the background taken from an annulus of 9 to 13 arcmin radius from the cluster centre around each proto-cluster. The colours mark different redshift intervals from $z = 1.5$ to 4.0 with a width of $\Delta z = 0.5$.

is a clear correlation, the estimated background variance can be reduced compared to that considered in the previous section if we add information about the local background.

We estimated the background in an annulus around the detection aperture with radii of 9 and 13 arcmin. The radii of proto-clusters for masses of 10^{14} and $10^{15} M_{\odot}$ for different redshifts can be found in Table 7. For $z = 1.5$ we find $r_{\text{pc}} \sim 4.4\text{--}9.6$ arcmin and for $z = 4$ we get $r_{\text{pc}} \sim 3.4\text{--}7.4$ arcmin for this mass range. The background region is thus outside the proto-clusters. The correlation of the background in the aperture and the annulus, shown in Fig. 14, helps us in the following way. A measurement of the background in the annulus can be used to normalise the background in the aperture region. The residual aperture background will then be smaller than the variance without this normalisation. The relevant residual background variance is then the variance of the ratio of the galaxy density inside the aperture to the galaxy density in the background annulus. It is the scatter of the relation shown in Fig. 14.

In Fig. 15 the rms of the density fluctuations is shown as a function of aperture radius and redshift in an equivalent way to the results in Fig. 11. Thus, the results can be directly compared in these figures, and we note, in most cases, an improvement of almost a factor of 2. The improvement is due to the fact that the variations in the projected galaxy density background are caused only to a minor degree by Poisson noise but mostly by large-scale structure, which can, for example, be well described by clustering statistics like the two-point correlation function. Therefore we obtain a better estimate of the local background if we take a measurement in its immediate neighbourhood.

This has, of course, an effect on the detection significance, which is illustrated in Fig. 16. We note that now more proto-clusters reach a detection significance of 2σ , and a large fraction of detections has significances above 1σ , which was not the case in the previous section.

Figure 17 (top) summarises the results on the detection significances. This figure can be compared to Fig. 13, with a change in the radius units to arcmin. We note that the local background assessment can provide a significant improvement. For real observations, the method can be tested by studying the spatial correlation of the galaxy density fluctuations in the field outside the proto-clusters and by using these results to determine the

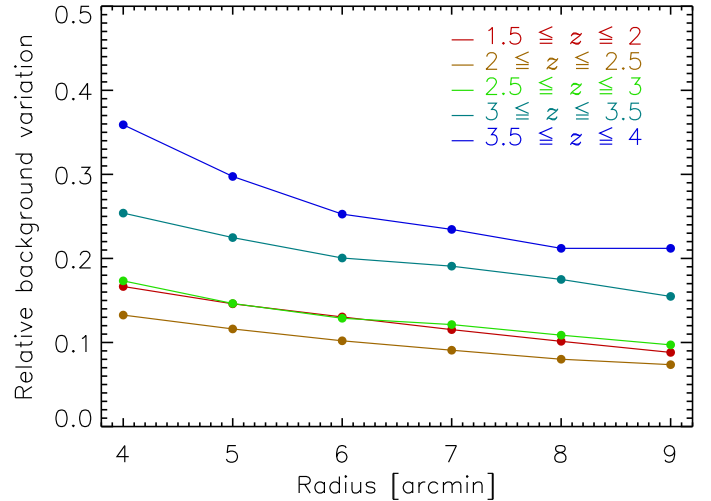


Fig. 15. Relative (rms) variations of the projected background galaxy densities (solid lines) with a local assessment as a function of aperture radius for five different redshift intervals.

rms of the background fluctuations. This can actually be applied in most detection methods, for example, for matched-filter detections. In practice, the filter would, for example, include the background ring with a negative weighting.

The bottom panel of Fig. 17 shows, analogously to Fig. 13 the mean detection significance as a function of the photometric redshift completeness limit for five different redshift intervals. The solid lines are for an aperture radius of $0.5 r_{\text{pc}}$ and the dashed lines for $1 r_{\text{pc}}$. We note that the results for the redshift intervals Δz (30%) and Δz (40%) have improved in comparison to the other redshift intervals, but Δz (50%) is still a good choice for the detection.

The practical meaning of the significance of a detection algorithm depends also on the rareness of the objects to be detected. If a large sky area has to be inspected to find an object, one needs a high detection significance threshold to keep the detected samples reasonably pure. In our case, we will find below that proto-clusters are quite abundant in projection on the sky. Therefore, we can still obtain a valuable proto-cluster candidate sample with a low significance threshold. The discussion section provides further details on this point.

In summary, we conclude that the best strategy for the detection of proto-clusters is to use an aperture radius smaller than r_{pc} and a local background assessment. Also, a redshift range given by Δz (50%) is quite optimal, but an exploration of higher completeness limits is often not much worse. A more sophisticated detection algorithm will, in this respect, anyway, include a probability distributions of photometric redshifts.

8. Mass-richness relation

For galaxy clusters, the richness, the number of member galaxies inside a given radius and magnitude limit can be used as a proxy for the cluster mass (e.g. Andreon & Hurn 2010; Rykoff et al. 2014; Castignani & Benoist 2016). Therefore, we test in this section how tight the mass-richness relation is for the proto-clusters in the simulations. Here, we inspect the intrinsic relation, including all known proto-cluster members in the simulations (with magnitude limits defined in Sect. 3) and not only the ones that would be detected with a certain prescription.

The mass-richness relation was determined for the MAMBO, GAEA H, and GAEA F samples of proto-clusters.

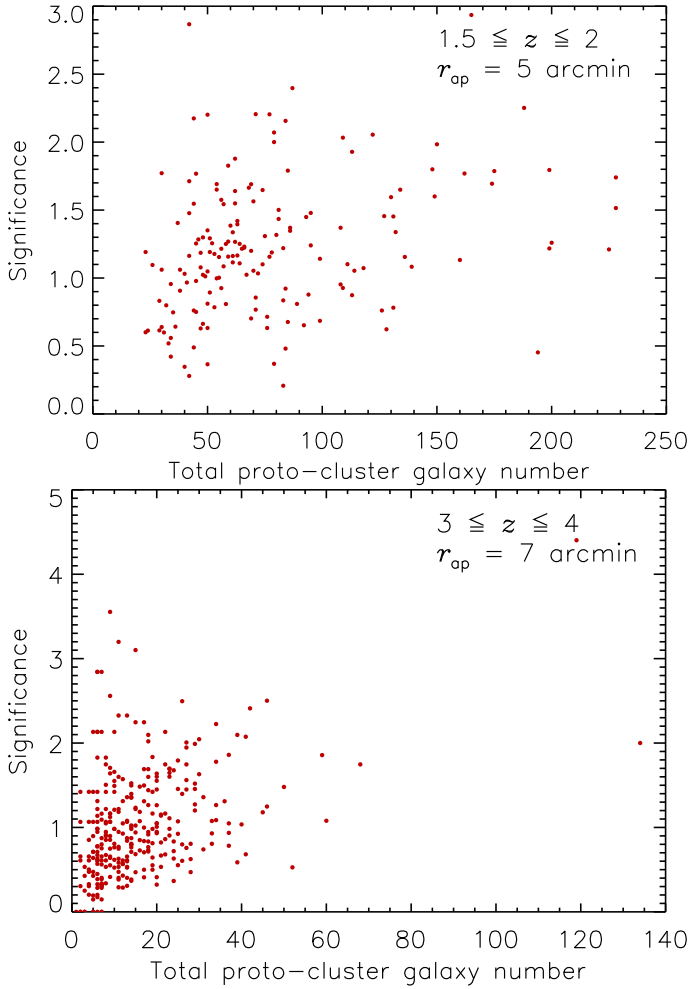


Fig. 16. Detection significance for proto-clusters for the redshift range $z = 1.5\text{--}2$ (top) and $z = 3\text{--}4$ (bottom). For the galaxy selection a redshift interval corresponding to $\Delta z_{(50\%)}$ was used. Top: A detection aperture of 5 arcmin was applied with a local background assessment. Bottom: An aperture of 7 arcmin was used.

Here, the mass of the system is that of the descendent cluster at $z = 0$ inside r_{200} since we attribute all the mass of the descendent cluster to the proto-cluster at any redshift. We binned the proto-clusters into subsamples of redshift bins with a width of $\Delta z = 0.25$ starting at $z = 1.5$ and an extra bin for $z = 1.5\text{--}2$. This leaves several hundred proto-clusters per bin for the GAEA samples and an order of a hundred for MAMBO.

The results are shown in Fig. 18 for the GAEA H simulations. The results from the other lightcones look similar. We find a few proto-clusters with very low member numbers well outside the variance of the number counts. They were identified as artefacts due to some common problems in the production of the lightcones. We excluded them from our study by a cut, which removes all cases with a negative 3σ deviation from the mean relation. The relative scatter that we observe in the relations, as shown for some examples in Fig. 18, is typically around 40% and decreases with richness to about 20%. This scatter is distinctly different from Poisson uncertainties and usually significantly larger. We provide some further illustrations of this fact towards the end of this Section.

The distribution of the number counts in Fig. 18 and all other relations studied is highly suggestive of a linear relation in logarithmic space. Therefore, we fitted the distribution by a relation

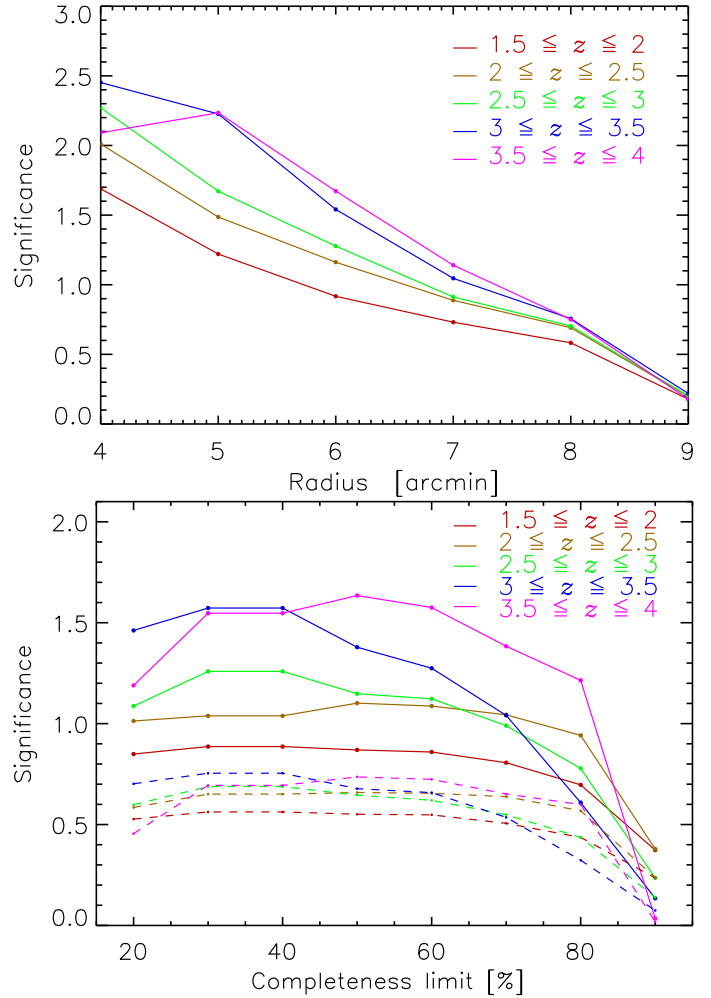


Fig. 17. Mean detection significance of proto-clusters as a function of aperture radius (top) and as a function of the completeness limit (bottom). The set-up is similar to Fig. 13, but with a local background and an aperture radius in arcminutes. Top: A redshift interval of $\Delta z_{(50\%)}$ was used. Bottom: Values of the completeness limit, Δz ($X\%$), in the range of 30–90% were used. The solid lines show the results for an aperture radius of $0.5 r_{pc}$ and the dashed lines show the results for $1 r_{pc}$.

of the form:

$$N_g(M, z) = N_{g0}(z) \left(\frac{M}{10^{14} M_\odot} \right)^\alpha. \quad (4)$$

The fits are shown in the figures as dashed lines. The fit results for the normalisation as a function of redshift are shown in the top panel of Fig. 19 for the three sets of simulations. The results are encouragingly similar. We fitted this mass-richness relation normalisation as a function of redshift with a third-order polynomial expression. The lines in Fig. 19 show the fitted functions, and the resulting parameters are given in Table 8.

The slope of the fitted mass-richness relation for the different redshift shells is shown in the bottom panel of Fig. 19. The values are about 0.9 or a little higher. Table 8 also shows the mean slopes averaged over the redshift intervals. This is in line with the observational finding that the efficiency of galaxy formation decreases with halo mass in the mass range above $10^{12} M_\odot$ (e.g. Behroozi et al. 2013; Kravtsov et al. 2018). The simulations attempted to reproduce this empirical finding. It is also good to observe that the mass-richness relations are similar

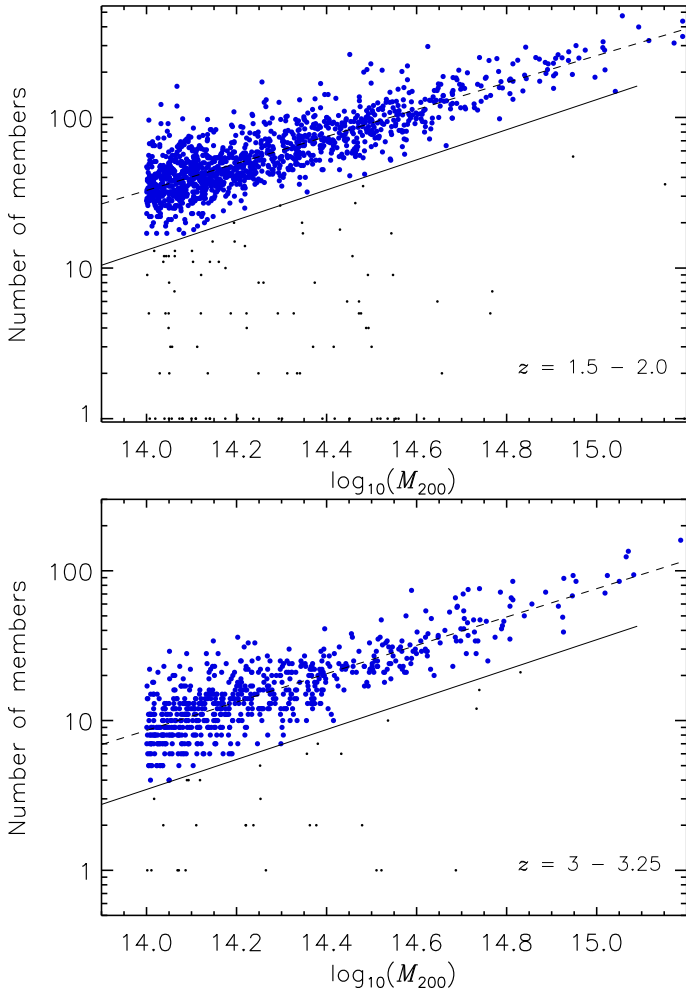


Fig. 18. Mass-richness relation for GAEA H for the redshift range $z = 1.5$ – 2 (top) and $z = 3$ – 3.25 (bottom). The small black dots show all proto-clusters, and the blue dots are the ones above the cut, which is shown as a solid black line. The linear regression fit to the relation is shown by a dashed line. M_{200} is in units of M_{\odot} .

in all three approaches of painting galaxy evolution onto the cosmological simulations. The small deviations are also due to the limited statistics.

We also studied the variance of the richness in the mass-richness relations because we wanted to see if we could approximate this quantity with simple Poisson statistics. We used the GAEA sample here for better statistics. We divided the sample into redshift bins with the division described above. Each of the eleven redshift bins was subdivided into several mass bins, with at least 50 members per interval. This left us with 12–25 mass bins per redshift bin. For each subsample, we determined the root mean square (RMS) deviation between the mean number of galaxies in the bin and the actual counts. The results are plotted for the GAEA H sample in Fig. 20. We compare the results for the scatter in the galaxy counts with the expectation for Poisson statistics. The relative scatter for the galaxy counts in the simulations is fairly constant as a function of the mean galaxy number: it decreases from about 0.4 to 0.25 over the range of galaxy counts from 3 to 200. This is distinctly different from the Poisson distribution, where the relative RMS changes by a factor of about 8. This deviation from Poisson statistics is observed in the simulations. At least here, the formation of galaxies in the overdense region of the proto-cluster is not a Poisson point pro-

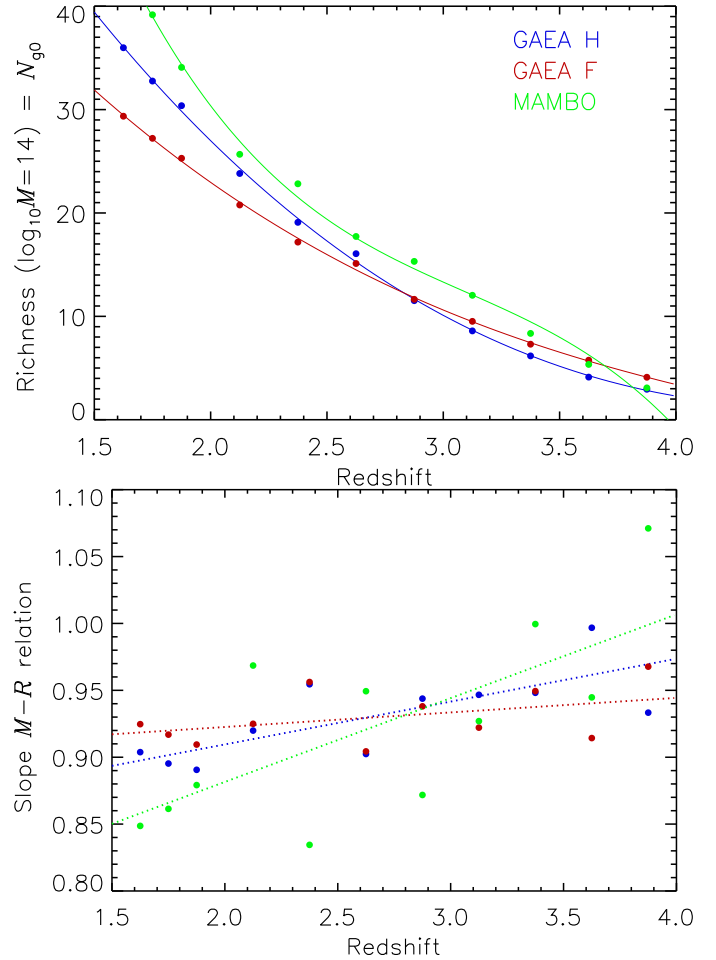


Fig. 19. Top: Normalisation, N_{g0} , for the mass-richness relation as a function of redshift. The lines show the third-order polynomial approximation. The normalisation is given for $\log_{10} M = 14$, with M in units of solar mass. Bottom: Slope parameter, α , of the mass-richness relation as a function of redshift. The dotted lines show fits of linear relations. The colours have the same meaning as above.

cess. It is not immediately clear if this is purely a result of the process of ‘painting galaxies onto the N -body simulations’ or if we should expect a similar behaviour in nature. But this exercise gives at least a warning that we should not use simple Poisson statistics for the uncertainty of the mass-richness relation unless we have shown that it works for observational data.

To further illustrate this difference between the variance of the number counts for a given mass and a Poisson distribution, we study the distribution of the richness for a given mass in two narrow mass intervals as shown in Fig. 21 for the redshift range $z = 1.5$ – 2 . The mean masses in the intervals are $\sim 1 \times 10^{14} M_{\odot}$ and $\sim 4.5 \times 10^{14} M_{\odot}$. To remove the scatter in the number counts due to the width of the mass interval, we normalise the number of members of each proto-cluster by a correction factor \bar{m}/m_{pc} . If for example the mass of a proto-cluster in the first bin is $1.15 \times 10^{14} M_{\odot}$, we multiply the number counts by a factor of $1/1.15$. We note that in both cases, the observed distribution is wider than the Poisson prediction, and this difference is much more pronounced in the higher mass bin.

Overall, we find mass-richness relations that are close to linear with logarithmic slopes close to 0.9. Again, we emphasise that these are the relations found in the simulations with known galaxy memberships.

Table 7. Statistics concerning the proto-cluster abundance and sky coverage as a function of redshift.

Redshift	Volume (deg ⁻²)	$N(>M_{14})$	$N(>2M_{14})$	$N(>5M_{14})$	$N(>M_{15})$	rad(M_{14}) (arcmin)	rad(M_{15}) (arcmin)	Area (deg ²)
0.50	0.68	5.47	1.64	0.21	0.02	6.528	14.064	0.0372
1.00	2.98	23.82	7.15	0.89	0.11	5.079	10.943	0.0225
1.50	4.78	38.21	11.46	1.43	0.17	4.447	9.581	0.0173
2.00	5.67	45.38	13.62	1.70	0.20	4.084	8.799	0.0146
2.50	5.98	47.86	14.36	1.79	0.22	3.842	8.278	0.0129
3.00	5.97	47.78	14.33	1.79	0.22	3.665	7.896	0.0117
3.50	5.80	46.43	13.93	1.74	0.21	3.527	7.599	0.0109
4.00	5.56	44.48	13.34	1.67	0.20	3.415	7.358	0.0102
4.50	5.29	42.29	12.69	1.59	0.19	3.322	7.157	0.0096
5.00	5.01	40.07	12.02	1.50	0.18	3.243	6.986	0.0092
5.50	4.74	37.91	11.37	1.42	0.17	3.174	6.839	0.0088
6.00	4.48	35.86	10.76	1.34	0.16	3.115	6.712	0.0085
6.50	4.24	33.92	10.18	1.27	0.15	3.064	6.601	0.0082
7.00	4.02	32.13	9.64	1.20	0.14	3.019	6.505	0.0080
7.50	3.81	30.45	9.13	1.14	0.14	2.981	6.422	0.0078
8.00	3.62	28.92	8.68	1.08	0.13	2.948	6.350	0.0076
8.50	3.43	27.46	8.24	1.03	0.12	2.919	6.290	0.0074
9.00	3.27	26.16	7.85	0.98	0.12	2.896	6.239	0.0073

Notes. Column (1) gives the upper bound of the redshift interval with width $\Delta z = 0.5$, Column (2) shows the volume per deg² in units of 10^6 Mpc^3 , Columns (3–6) provide the number of proto-clusters in this volume with mass $>10^{14}$, 2×10^{14} , 5×10^{14} and $10^{15} M_{\odot}$, respectively, Columns (7,8) give the physical radii of proto-clusters with masses of 10^{14} and $10^{15} M_{\odot}$, and Column (9) shows the sky area covered by a proto-cluster with a mass $\geq 10^{14} M_{\odot}$.

9. Proto-cluster abundance and sky coverage

The abundance of proto-clusters is determined by their definition, which in our study is given by the statement that they evolve into present-day galaxy clusters with masses $M_{200} \geq 10^{14} M_{\odot}$. Practically, we defined proto-clusters above as all galaxies that will end up in the descendant cluster at $z = 0$ inside r_{200} . Similarly, we associate all matter that will finally be assembled in the zero redshift cluster as belonging to the proto-cluster. This is exactly the mass contained inside r_{pc} in the top-hat overdensity model. Thus, in this context, the proto-cluster has the same mass as the descendant cluster at all times. Therefore, the density of proto-clusters for a certain mass limit in comoving coordinates is that of the present-day clusters above that mass limit. The cumulative present-day cluster mass function, $n(>M)$, can be obtained from observations, for example, from the cluster abundance in the REFLEX survey of X-ray luminous galaxy clusters (Böhringer et al. 2013, 2014), one of the best defined and comprehensive galaxy cluster samples in the nearby Universe. The cumulative mass function can be approximated by the following function:

$$n(>M) = \alpha \left(\frac{M_{14}}{2} \right)^{-\beta} \exp \left(\frac{-M_{14}^{\delta}}{\gamma} \right), \quad (5)$$

where M_{14} is the proto-cluster mass in units of $10^{14} M_{\odot}$ for a fiducial radius of r_{200} , $\alpha = 1.237 \times 10^{-5} \text{ Mpc}^{-3}$, $\beta = 0.907$, $\gamma = 0.961$, and $\delta = 0.625$ (Böhringer et al. 2017). For the calculation of the comoving volume as a function of redshift, we used the reference cosmology. Table 7 shows the results for different lower mass limits for volumes for which proto-cluster counts were integrated over redshift intervals of $\Delta z = 0.5$. The redshifts listed in the Table, give the upper bound of the redshift interval. The Table also provides values for the proto-cluster radii calculated by means of Eq. (2) in units of arcmin on the sky for

Table 8. Fit parameters of the polynomial expression for the normalisation, $N_{\text{go}}(z) = a + b z + c z^2 + d z^3$, of the mass-richness relation as a function of redshift for the three proto-cluster samples. We also show the slope, α , averaged over the redshift intervals.

Sample	a	b	c	d	$\langle \alpha \rangle$
GAEA H	95.512	-47.695	7.339	-0.3104	0.93
GAEA F	73.268	-36.778	6.577	-0.4452	0.93
MAMBO	206.466	-168.640	51.386	-5.5446	0.92

proto-clusters with masses of $10^{14} M_{\odot}$ and $10^{15} M_{\odot}$ and sky surface areas of proto-clusters with masses $\geq 10^{14} M_{\odot}$.

Based on the proto-cluster abundances described above, we can study how well we can separate proto-clusters in the sky if we have only photometric redshifts. We look at the fraction of the sky covered by proto-clusters within a redshift slice, which corresponds to the uncertainty of photometric redshifts. For the study, we select all the proto-clusters and their member galaxies from the MAMBO lightcone in redshift intervals given by the photometric redshift accuracy for a 50% completeness limit and show how they are projected onto the sky. Figure 22 shows proto-clusters and member galaxies selected from redshifts around $z = 2$ with a redshift range of $\Delta z_{(50\%)}$. The core regions ($r \leq 0.5 r_{\text{pc}}$), which will stick out due to their higher density contrast, are highlighted in the plot.

In total, there are 132 proto-clusters; in this exercise, we also included proto-clusters that are not completely contained in the field of view of the MAMBO lightcone.

We note that the sky is indeed densely covered by the proto-clusters. By just integrating the surface area of each proto-cluster (inside r_{pc}), taking into account that some proto-cluster have some of their areas outside the lightcone, we find that the sum

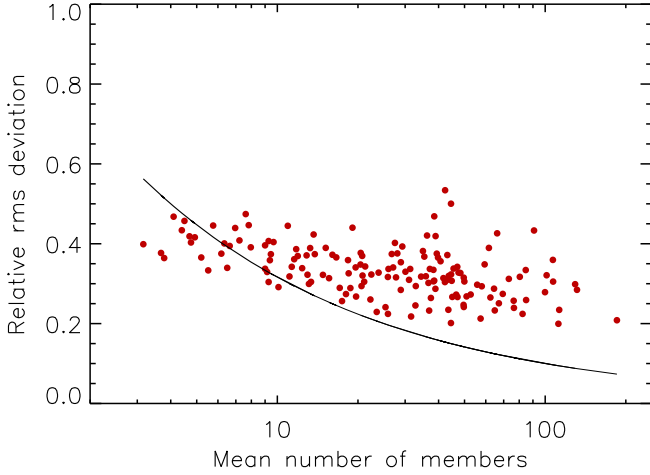


Fig. 20. Relative rms deviation of the galaxy number counts as a function of the mean number of galaxies for the proto-clusters in each mass-redshift bin (data points). We also show the expectation for the scatter if it would be governed by Poisson statistics, which is shown by the solid line.

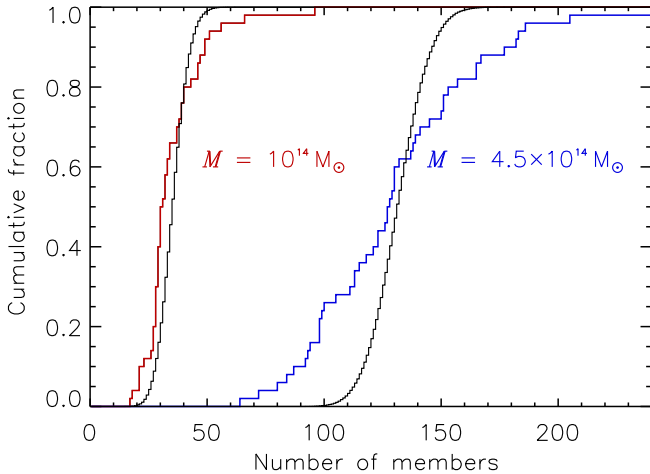


Fig. 21. Comparison of the distribution of the galaxy membership counts (richness) for given mass and the prediction for Poissonian scatter (black curves). The mean masses in the interval are $10^{14} M_{\odot}$ (red curve) and $4.5 \times 10^{14} M_{\odot}$ (blue curve), respectively. The curves show cumulative histograms normalised to unity. The data are taken from the GAEA F lightcone in the redshift interval $z = 1.5-2$.

of all proto-cluster areas is 91%, while the fraction of covered sky is about 59%, due to overlaps. We have seen in Sects. 6 and 7, however, that proto-clusters are detected with sufficient significance around 2σ only at radii smaller than the proto-cluster radius and in Sect. 5, we showed that typically half the member galaxies are found inside $0.5 r_{\text{pc}}$ where they have a 2.7 times higher projected density than in the annulus around. Therefore, to give an impression of the sky coverage of the recognisable core regions of the proto-clusters, we also quote the results for proto-cluster radii of $0.5 r_{\text{pc}}$, which is a total fractional area of 23% and a fraction of the sky covered of 21%. Figure 23 illustrates the case for both proto-cluster radii. We note that in some regions, the proto-cluster cores overlap, which is the result of clustering in large-scale structures.

We can, however, look at this from a more practical point of view. If we relax our ambition to detect all proto-clusters with a final mass of $10^{14} M_{\odot}$ and look, for example, only at

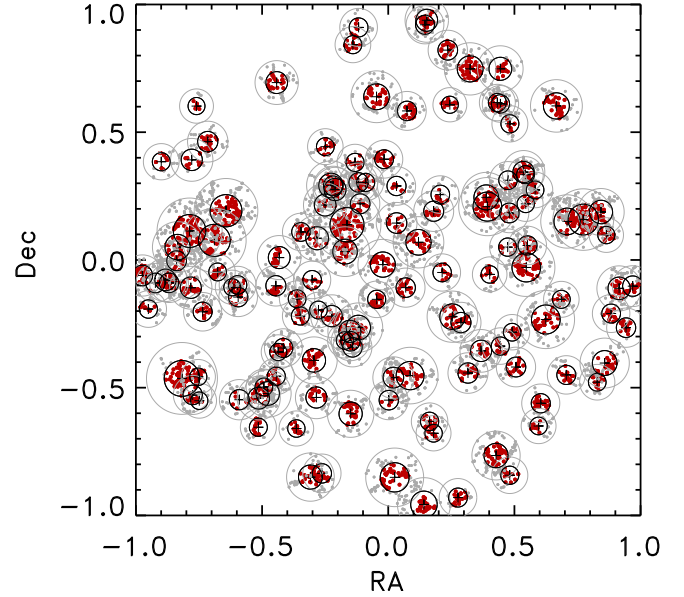


Fig. 22. Proto-clusters and their galaxies in a redshift slice around $z = 2$ with a redshift range according to the photometric redshift accuracy limits for 50% completeness and $M_{\text{pc}} \geq 10^{14} M_{\odot}$. The dense core regions inside $r \leq 0.5 r_{\text{pc}}$ are shown by black circles and red dots for the galaxies, while the outer regions are shown in grey.

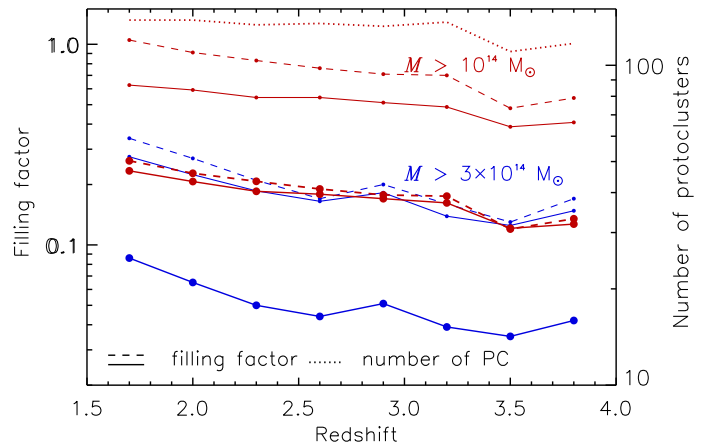


Fig. 23. Number of MAMBO proto-clusters and the sky-filling factor in a $\Delta_z (50\%)$ redshift interval as a function of redshift and proto-cluster mass limit. The numbers are only shown for the sample with $M_{\text{pc}} \geq 10^{14} M_{\odot}$ as a dotted line. Solid lines show the fraction of the sky covered by the proto-clusters, and the dashed lines display the combined sky area of the proto-clusters. The thick lines with larger data points show the results for the core regions, while the thin lines show those for the region out to r_{pc} . For the core regions of the $M_{\text{pc}} \geq 3 \times 10^{14} M_{\odot}$ proto-clusters, the dashed and solid lines overlap.

proto-clusters above a mass limit of $3 \times 10^{14} M_{\odot}$, the situation looks much better, as shown in Fig. 24. If we increase the mass limit even further, for example, to $8 \times 10^{14} M_{\odot}$ we find only none to three proto-clusters per $\Delta_z (50\%)$ in the MAMBO lightcone, with filling factors of a few per cent. Only in the highest redshift bin, are there four proto-clusters with a filling factor of 7%. These larger mass objects correspond probably better to the typical proto-clusters that have been identified in optical surveys. If we had included smaller systems ($< 10^{14} M_{\odot}$), the picture would have looked worse. Thus, the whole problem is that of detecting massive proto-clusters in the presence of large-scale

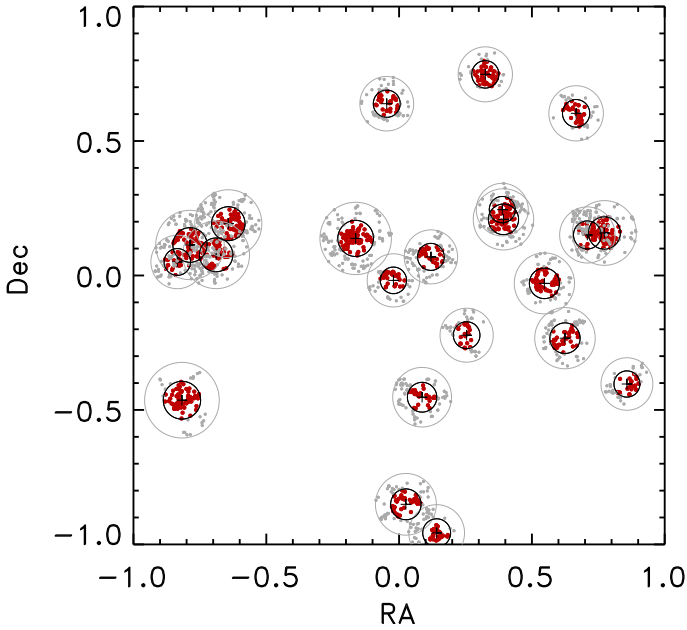


Fig. 24. Proto-clusters and their galaxies in a redshift slice around $z = 2$ as in Fig. 22 but with a higher mass limit of the proto-clusters of $M_{\text{pc}} = 3 \times 10^{14} M_{\odot}$. Symbols and colours have the same meaning as in Fig. 22.

structures in the background, a problem that we have already encountered in section 6, which has already been taken into account in the calculation of the detection significance. It therefore appears that targeting proto-clusters with a mass limit of $10^{14} M_{\odot}$ may be ambitious. In Fig. 25 we also show the sky coverage for $M > 10^{14} M_{\odot}$ proto-clusters around redshift $z \sim 3.5$. The sky coverage is less dramatic than in the low redshift case. But we note in all three pictures the substantial clumping of the proto-clusters. This may sometimes cause difficulties in clearly separate structures, which are identified with separate proto-clusters in the simulations.

In Fig. 23, we summarise the results for our target redshift range at eight redshift values. We show the sky coverage of the proto-clusters in the $\Delta_z(50\%)$ redshift slices and also the sum of their sky area. The sky coverage factors are decreasing with redshift. Since only the centres of the proto-clusters will be detectable and their sky-filling factors never exceed 30% even for the worst case, proto-clusters should, in general, be separable from detection algorithms, except for the regions of proto-superclusters.

10. Discussion

The previous sections have shown that the approximate analytic concept for the prediction of the proto-cluster properties such as size, density contrast, and abundance, developed above, can provide a robust guideline to explore proto-clusters in the Euclid Wide Survey. In particular, the prediction for the proto-cluster radius gives a good approach to defining a fiducial radius for a proto-cluster. In addition, we have obtained further useful descriptions, such as the typical proto-cluster profiles and their variations by means of the simulations. We note that some of the presented properties are intrinsic in the sense that they are described on the basis of the information given by the simulations. This includes the proto-cluster radii, profiles, abundances, and mass-richness relation. In contrast the detection significances are observed properties, and in their case the information

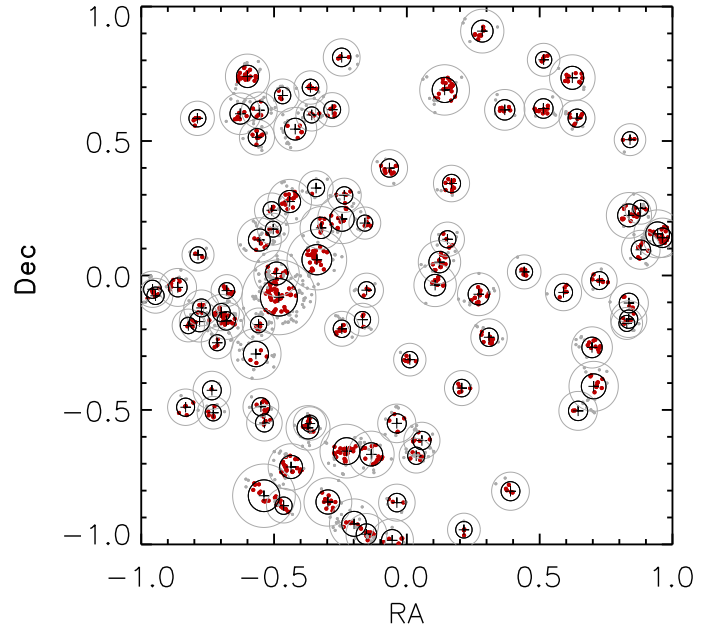


Fig. 25. Proto-clusters with $M_{\text{pc}} \geq 10^{14} M_{\odot}$ and their galaxies in a redshift slice around $z = 3.5$ with a redshift range according to the photometric redshift accuracy limits for 50% completeness. Symbols and colours have the same meaning as in Fig. 22.

from the simulations that is used in the same as would be available to an observer.

As an application of the information given in Table 7, we calculate the expected volumes of proto-clusters and compare them to some observed examples in the literature. From the radii given in arcmin, we calculate the comoving volumes for proto-clusters with a descendent mass of $10^{15} M_{\odot}$ at redshifts 2, 3, 4, and 5 with results of about 9700, 13 000, 15 000, and 16 000 Mpc^3 , respectively, and compare them to the volumes assigned to some prominent proto-clusters by Casey (2016). The quoted volumes for COSMOS ($z = 2.10$) and ($z = 2.47$) of about 15 000 Mpc^3 are larger than what is expected for the most massive proto-clusters, and one might conclude that we are looking at two or more closely neighbouring structures, as we see them in Figs. 22, 24, and 25. The combined volume of three fields in SSA22 ($z = 3.09$) with a value of 21 000 Mpc^3 is consistent with the expectation for massive proto-clusters if we take the three fields as different systems. The value of 20 000 Mpc^3 for HDF 850.1 ($z = 5.18$) is a bit high but could still be explained by one very massive proto-cluster if the boundary was considered slightly too generous. For MRC1138–256 ($z = 2.16$) and AzTEC-3 ($z = 5.18$) the observed values are comparably small, and either the volumes assigned concern only part of the proto-cluster, or these proto-clusters have descendants with a mass much lower than $10^{15} M_{\odot}$. In a recent CO survey around MRC1138–256 (Jin et al. 2021), a large overdensity of CO emitters was found as part of a filamentary structure with an extent of about 120 comoving Mpc. The extent of 38 Mpc in physical scale is too large for one proto-cluster and a structure of this size should break up to form several virialised units. But within this filament, MRC1138–256 could indeed be a larger proto-cluster than mentioned above.

With these tools, we can prepare the strategy for the detection of proto-clusters in the Euclid Survey and the assessment of their properties. They could, in particular, provide a good

orientation for the parameter selection for more sophisticated detection algorithms than the aperture counts used here.

In Sects. 6 and 7, we reported the moderate detection significance for proto-clusters, which can be achieved with the Euclid Survey with the given photometric redshift uncertainties based on broadband photometry. This will make it difficult, in general, to characterise proto-clusters quantitatively without further follow-up observations. On the other hand, many proto-clusters will be detectable with sufficient significance ($\sim 2\sigma$) to obtain a large number of highly likely candidates. Since the proto-clusters have a large sky-filling factor, as shown in Sect. 10, one can actually afford a low detection significance threshold. The following example may best illustrate this. If we study a contiguous sky region of 200 times the area of a typical proto-cluster and the region contains 100 such systems, corresponding to a filling factor of 50%, a 2σ detection threshold would lead to an average false detection rate of 5 contaminating events (since only the positive deviations of the distribution count), a false detection rate of 5%, which could be tolerable for statistical studies.

What we have described here is a worst-case scenario using a very simple detection aperture. Several sophisticated detection algorithms have been developed for the detection of proto-clusters in the Euclid Survey. Using a particular aperture and redshift interval exploits only part of the information given by the observations. A good detection algorithm will make use of the complete available information, including galaxies at all relevant radii and the complete probability distributions of the photometric redshifts with optimal weights. Therefore, we can easily expect a significant increase in the detection efficiency with the proper use of dedicated algorithms. One has to be careful, however, when proto-cluster shapes are assumed in the algorithms about the possible introduction of selection effects. The examples of unusual proto-cluster profiles found in Sect. 5 are, in this respect, interesting test cases.

We can illustrate the unique opportunity offered by *Euclid*, also with the following example. We focus on the most massive proto-clusters with $M \geq 8 \times 10^{14} M_{\odot}$ mentioned above with a typical sky-filling factor of $\sim 5\%$. A dedicated algorithm is expected to improve the detection significance from ~ 2 to at least $\sim 3\sigma$. This would correspond to false detections with a sky-filling factor of 0.15%. Thus, we would obtain a ratio of about 34 true to one false detection. In total, we can expect of the order of 40 000 proto-clusters with a mass limit of $M \geq 8 \times 10^{14} M_{\odot}$ in the complete Euclid Wide Survey in the redshift range $z = 1.5-4$. This will definitely provide a relatively pure, interesting, and unique proto-cluster candidate sample.

A sweet spot for the detection of proto-clusters in the considered redshift range is $z \sim 3$, where the density contrast turns out to be the highest. This is an interesting region where Ly-break galaxy surveys revealed the first proto-clusters. With a further boost of significance by means of dedicated detection algorithms, the data should be sufficient to estimate selection functions and perform population statistics. What will be particularly interesting is the unprecedentedly large survey volume of the Euclid Wide Survey, which will turn up the most extreme and rare objects that have not been observed in the available much smaller survey areas.

The present study also leads to the question: how can the contrast above the background be improved? Narrowband surveys will, of course, lead to more precise photometric redshifts and thus provide an improvement. We can take a look at two examples. The survey of Yamada et al. (2012) in the SA22 field at $z \sim 3$ with a narrowband filter $\Delta\lambda = 77 \text{ \AA}$ provides an accuracy of about $\Delta z = 0.063$. For the COSMOS survey Chiang et al.

(2014) obtained $\Delta z = 0.025 (1 + z)$. Compared to the numbers for the *Euclid* data given above, this is an improvement of about a factor of two, which is, of course, very helpful, but it does not improve the situation so much that it allows a precise assessment of the proto-cluster properties. Thus, spectroscopic follow-up observations will be mandatory for an accurate study of the most interesting proto-clusters.

11. Summary and conclusions

We have provided an overview of how galaxy proto-clusters in the redshift range $z = 1.5-4$ are expected to appear in the Euclid Wide Survey in order to assist in the preparation of the survey analysis. The paper reports results on the following studies of proto-cluster properties.

- A practical estimate of the proto-cluster radius was obtained on the basis of an analytical model for the evolution of a homogeneous top-hat overdensity. A comparison with cosmological simulations showed that about 80% of the members of the descendent cluster (inside r_{200}) are contained within this radius with contamination of about 5–10%. We provided an analytic approximation of the evolution of the proto-cluster radius with redshift (Sect. 4).
- The mean radial galaxy distribution of proto-clusters can be well described with a cored profile with a specified inner and outer logarithmic slope. We provide best-fit formulae for the mean three-dimensional and projected profiles (Sect. 5). Individual proto-clusters, however, show a large variety. About three-quarters of the proto-clusters have a pronounced dense core. However, proto-clusters with multiple substructure components can have profiles with a density maximum at larger radii.
- We investigated which density contrast in the galaxy distribution proto-clusters can be observed against the galaxy background. We assumed that photometric data from the Euclid Survey and broadband ground-based auxiliary data (at the time of the third data release) are available for this study to estimate photometric redshifts. We showed that the small overdensity of proto-clusters, the large redshift depth that has to be sampled, and the large-scale structure variance of the background lead to a moderate detection significance for simple detections in circular apertures. With a local background assessment and apertures focused on the inner regions of the proto-clusters, significances of the order of 2σ can be reached. Because proto-clusters are not very rare, useful samples of highly likely proto-cluster candidates can be obtained with such detection thresholds.
- The cluster richness (i.e. the number of galaxies that are members of the proto-cluster) was found to be tightly correlated with the mass of the descendent cluster. We found a relation with a logarithmic slope around of 0.9, which may reflect the decreasing galaxy formation efficiency with increasing halo mass for the most massive halos. We provided analytical approximations to the mass-richness relation as a function of mass and redshift for the GAEA and MAMBO simulations (Sect. 8).
- The abundance of proto-clusters is determined by their definition of being the precursors of galaxy clusters at redshift zero with masses, $M_{200} \geq 10^{14} M_{\odot}$. The matter fraction in such clusters was, for example, determined in Böhlinger et al. (2017) to be about 4%, and the comoving number density can be obtained from this result. By knowing the abundance and sizes of the objects, one can calculate how densely the proto-clusters fill the sky for a redshift range comparable to

the uncertainties of the photometric redshifts. We found that the area that is filled is quite large, and proto-clusters with the quoted mass limit start to overlap significantly.

There are several efforts in the Euclid Consortium to devise optimised detection algorithms for the search of proto-clusters in the Euclid Survey. They will use more sophisticated techniques, which also involve probability distributions for the photometric galaxy redshifts. This will provide the leverage to improve the detection efficiency over what has been described here. With the large sky area explored by *Euclid*, an enormously large database of likely proto-clusters and their properties will be provided, which can bring proto-cluster studies and follow-up observations to a new level.

Acknowledgements. HB thanks the Munich Excellence Cluster ‘Origins’ for support. GC acknowledges support by the Deutsches Luft- und Raumfahrt Zentrum through grant no. 50 OR 2204. HD acknowledges financial support from the Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación (AEI-MCINN) under grant (La evolución de los cúmulos de galaxias desde el amanecer hasta el mediodía cósmico) with reference (PID2019-105776GB-I00/DOI:10.13039/501100011033) and del Ministerio de Ciencia, Innovación y Universidades (MCIU/AEI) under grant (Construcción de cúmulos de galaxias en formación a través de la formación estelar ocurecida por el polvo) and the European Regional Development Fund (ERDF) with reference (PID2022-143243NB-I00/DOI:10.13039/501100011033). LM acknowledges the financial contribution from the grant PRIN-MUR 2022 20227RNLY3 “The concordance cosmological model: stress-tests with galaxy clusters” supported by Next Generation EU and from the grants ASI n.2018-23-HH.0 and n. 2024-10-HH.0 “Attività scientifiche per la missione Euclid – fase E” The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Agenzia Spaziale Italiana, the Austrian Forschungsförderungsgesellschaft funded through BMK, the Belgian Science Policy, the Canadian Euclid Consortium, the Deutsches Zentrum für Luft- und Raumfahrt, the DTU Space and the Niels Bohr Institute in Denmark, the French Centre National d’Etudes Spatiales, the Fundação para a Ciência e a Tecnologia, the Hungarian Academy of Sciences, the Ministerio de Ciencia, Innovación y Universidades, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Research Council of Finland, 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. A complete and detailed list is available on the Euclid web site (www.euclid-ec.org).

References

- Alberts, S., & Noble, A. 2022, *Universe*, 8, 554
- Andreon, S., & Hurn, M. A. 2010, *MNRAS*, 404, 1922
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57
- Böhringer, H., Chon, G., Collins, C. A., et al. 2013, *A&A*, 555, A30
- Böhringer, H., Chon, G., & Collins, C. A. 2014, *A&A*, 570, A31
- Böhringer, H., Chon, G., & Fukugita, M. 2017, *A&A*, 608, A65
- Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, *ApJ*, 379, 440
- Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, *MNRAS*, 398, 1150
- Calvi, R., Castignani, G., & Dannerbauer, H. 2023, *A&A*, 678, A15
- Casey, C. M. 2016, *ApJ*, 824, 36
- Casey, C. M., Cooray, A., Capak, P., et al. 2015, *ApJ*, 808, L33
- Castignani, G., & Benoist, C. 2016, *A&A*, 595, A111
- Castignani, G., Chiaberge, M., Celotti, A., & Norman, C. 2014a, *ApJ*, 792, 113
- Castignani, G., Chiaberge, M., Celotti, A., Norman, C., & De Zotti, G. 2014b, *ApJ*, 792, 114
- Castro, T., Borgani, S., Dolag, K., et al. 2021, *MNRAS*, 500, 2316
- Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, *ApJ*, 779, 127
- Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2014, *ApJ*, 782, L3
- Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. 2017, *ApJ*, 844, L23
- Chon, G., Böhringer, H., & Zaroubi, S. 2015, *A&A*, 575, L14
- Clements, D. L., Braglia, F. G., Hyde, A. K., et al. 2014, *MNRAS*, 439, 1193
- Contini, E., De Lucia, G., Hatch, N., Borgani, S., & Kang, X. 2016, *MNRAS*, 456, 1924
- Cucciati, O., Zamorani, G., Lemaux, B. C., et al. 2014, *A&A*, 570, A16
- Cucciati, O., Lemaux, B. C., Zamorani, G., et al. 2018, *A&A*, 619, A49
- Daddi, E., Dannerbauer, H., Stern, D., et al. 2009, *ApJ*, 694, 1517
- Dannerbauer, H., Kurk, J. D., De Breuck, C., et al. 2014, *A&A*, 570, A55
- De Lucia, G., Tornatore, L., Frenk, C. S., et al. 2014, *MNRAS*, 445, 970
- De Lucia, G., Xie, L., Fontanot, F., & Hirschmann, M. 2020, *MNRAS*, 498, 3215
- Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, *MNRAS*, 456, 2486
- Djorgovski, S. G., Stern, D., Mahabal, A. A., & Brunner, R. 2003, *ApJ*, 596, 67
- Euclid Collaboration (Scaramella, R., et al.) 2022, *A&A*, 662, A112
- Euclid Collaboration (Cropper, M., et al.) 2024a, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202450996>
- Euclid Collaboration (Jahnke, K., et al.) 2024b, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202450786>
- Euclid Collaboration (Hormuth, F., et al.) 2024c, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202450345>
- Euclid Collaboration (Mellier, Y., et al.) 2024d, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202450810>
- Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., et al. 2002, *ApJ*, 573, 7
- Flores-Cacho, I., Pierini, D., Soucaill, G., et al. 2016, *A&A*, 585, A54
- Fontanot, F., De Lucia, G., Hirschmann, M., et al. 2017, *MNRAS*, 464, 3812
- Fontanot, F., De Lucia, G., Hirschmann, M., et al. 2020, *MNRAS*, 496, 3943
- Fontanot, F., Calabrò, A., Talia, M., et al. 2021, *MNRAS*, 504, 4481
- Francis, P. J., Woodgate, B. E., Warren, S. J., et al. 1996, *ApJ*, 457, 490
- Fu, H., Shankar, F., Ayromlou, M., et al. 2024, *MNRAS*, 532, 177
- Galametz, A., Stern, D., De Breuck, C., et al. 2012, *ApJ*, 749, 169
- García-Vergara, C., Hennawi, J. F., Barrientos, L. F., & Rix, H.-W. 2017, *ApJ*, 848, 7
- García-Vergara, C., Hennawi, J. F., Barrientos, L. F., & Arrigoni Battaia, F. 2019, *ApJ*, 886, 79
- García-Vergara, C., Rybak, M., Hodge, J., et al. 2022, *ApJ*, 927, 65
- Girelli, G., Pozzetti, L., Bolzonella, M., et al. 2020, *A&A*, 634, A135
- Gómez-Guijarro, C., Riechers, D. A., Pavesi, R., et al. 2019, *ApJ*, 872, 117
- Greenslade, J., Clements, D. L., Cheng, T., et al. 2018, *MNRAS*, 476, 3336
- Harikane, Y., Ouchi, M., Ono, Y., et al. 2019, *ApJ*, 883, 142
- Hatch, N. A., Kurk, J. D., Pentericci, L., et al. 2011a, *MNRAS*, 415, 2993
- Hatch, N. A., De Breuck, C., Galametz, A., et al. 2011b, *MNRAS*, 410, 1537
- Hennawi, J. F., Prochaska, J. X., Cantalupo, S., & Arrigoni-Battaia, F. 2015, *Science*, 348, 779
- Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, *MNRAS*, 451, 2663
- Higuchi, R., Ouchi, M., Ono, Y., et al. 2019, *ApJ*, 879, 28
- Hirschmann, M., De Lucia, G., & Fontanot, F. 2016, *MNRAS*, 461, 1760
- Ito, K., Tanaka, M., Valentino, F., et al. 2023, *ApJ*, 945, L9
- Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, *MNRAS*, 321, 372
- Jin, S., Dannerbauer, H., Emonts, B., et al. 2021, *A&A*, 652, A11
- Koyama, Y., Kodama, T., Tadaki, K.-I., et al. 2013, *MNRAS*, 428, 1551
- Kravtsov, A. V., Vikhlinin, A. A., & Meshcheryakov, A. V. 2018, *Astron. Lett.*, 44, 8
- Kuiper, E., Hatch, N. A., Venemans, B. P., et al. 2011, *MNRAS*, 417, 1088
- Kurk, J. D., Röttgering, H. J. A., Pentericci, L., et al. 2000, *A&A*, 358, L1
- Kurk, J. D., Pentericci, L., Röttgering, H. J. A., & Miley, G. K. 2004, *A&A*, 428, 793
- Le Fèvre, O., Deltorn, J. M., Crampton, D., & Dickinson, M. 1996, *ApJ*, 471, L11
- Lemaux, B. C., Le Floch, E., Le Fèvre, O., et al. 2014, *A&A*, 572, A90
- McConachie, I., Wilson, G., Forrest, B., et al. 2022, *ApJ*, 926, 37
- Miley, G., & De Breuck, C. 2008, *A&ARv*, 15, 67
- Miller, T. B., Chapman, S. C., Aravena, M., et al. 2018, *Nature*, 556, 469
- Muldrew, S. I., Hatch, N. A., & Cooke, E. A. 2015, *MNRAS*, 452, 2528
- Oteo, I., Ivison, R. J., Dunne, L., et al. 2018, *ApJ*, 856, 72
- Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2005, *ApJ*, 620, L1
- Overzier, R. A. 2016, *A&ARv*, 24, 14
- Pentericci, L., Kurk, J. D., Röttgering, H. J. A., et al. 2000, *A&A*, 361, L25
- Planck Collaboration XXVII. 2015, *A&A*, 582, A30
- Planck Collaboration XIII. 2016, *A&A*, 594, A13
- Press, W. H., & Schechter, P. 1974, *ApJ*, 187, 425
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, *ApJ*, 785, 104
- Sartoris, B., Biviano, A., Fedeli, C., et al. 2016, *MNRAS*, 459, 1764
- Schreiber, C., Elbaz, D., Pannella, M., et al. 2017, *A&A*, 602, A96
- Sheth, R. K., & Tormen, G. 1999, *MNRAS*, 308, 119
- Shimasaku, K., Ouchi, M., Okamura, S., et al. 2003, *ApJ*, 586, L111
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Steidel, C. C., Adelberger, K. L., Dickinson, M., et al. 1998, *ApJ*, 492, 428
- Strazzullo, V., Daddi, E., Gobat, R., et al. 2015, *A&A*, 576, L6
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, 724, 878
- Toshikawa, J., Kashikawa, N., Ota, K., et al. 2012, *ApJ*, 750, 137
- Tozzi, P., Gilli, R., Liu, A., et al. 2022, *A&A*, 667, A134
- Venemans, B. P., Kurk, J. D., Miley, G. K., et al. 2002, *ApJ*, 569, L11
- Venemans, B. P., Röttgering, H. J. A., Overzier, R. A., et al. 2004, *A&A*, 424, L17
- Venemans, B. P., McMahon, R. G., Warren, S. J., et al. 2007, *MNRAS*, 376, L76

- Vieira, J. D., Crawford, T. M., Switzer, E. R., et al. 2010, *ApJ*, 719, 763
- Wang, G. C. P., Hill, R., Chapman, S. C., et al. 2021, *MNRAS*, 508, 3754
- Wylezalek, D., Vernet, J., De Breuck, C., et al. 2013, *MNRAS*, 428, 3206
- Xie, L., De Lucia, G., Hirschmann, M., & Fontanot, F. 2020, *MNRAS*, 498, 4327
- Yamada, T., Nakamura, Y., Matsuda, Y., et al. 2012, *AJ*, 143, 79
- Zoldan, A., De Lucia, G., Xie, L., Fontanot, F., & Hirschmann, M. 2017, *MNRAS*, 465, 2236
-
- 1 Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
 - 2 Ludwig-Maximilians-University, Schellingstrasse 4, 80799 Munich, Germany
 - 3 Max-Planck-Institut für Physik, Boltzmannstr. 8, 85748 Garching, Germany
 - 4 INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
 - 5 Instituto de Astrofísica de Canarias (IAC); Departamento de Astrofísica, Universidad de La Laguna (ULL), 38200 La Laguna, Tenerife, Spain
 - 6 INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
 - 7 Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, Bd de l’Observatoire, CS 34229, 06304 Nice Cedex 4, France
 - 8 Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
 - 9 INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
 - 10 Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, 40126 Bologna, Italy
 - 11 School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK
 - 12 INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
 - 13 Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
 - 14 INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
 - 15 IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
 - 16 Institute of Physics, Laboratory for Galaxy Evolution, Ecole Polytechnique Fédérale de Lausanne, Observatoire de Sauverny, CH-1290 Versoix, Switzerland
 - 17 Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
 - 18 Zentrum für Astronomie, Universität Heidelberg, Philosophenweg 12, 69120 Heidelberg, Germany
 - 19 Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
 - 20 Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town 7535, South Africa
 - 21 Université Paris-Saclay, CNRS, Institut d’astrophysique spatiale, 91405 Orsay, France
 - 22 ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
 - 23 INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste TS, Italy
 - 24 SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste TS, Italy
 - 25 Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
 - 26 INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
 - 27 Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146 Genova, Italy
 - 28 INFN-Sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy,
 - 29 Department of Physics “E. Pancini”, University Federico II, Via Cinthia 6, 80126 Napoli, Italy
 - 30 INAF-Osservatorio Astronomico di Capodimonte, Via Moiarriello 16, 80131 Napoli, Italy
 - 31 INFN section of Naples, Via Cinthia 6, 80126 Napoli, Italy
 - 32 Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
 - 33 Faculdade de Ciências da Universidade do Porto, Rua do Campo de Alegre, 4150-007 Porto, Portugal
 - 34 Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
 - 35 INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
 - 36 INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
 - 37 Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
 - 38 Port d’Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
 - 39 Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
 - 40 Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
 - 41 Institut d’Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
 - 42 INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
 - 43 Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
 - 44 Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, 38204 San Cristóbal de La Laguna, Tenerife, Spain
 - 45 Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
 - 46 Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
 - 47 European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
 - 48 Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne F-69100, France
 - 49 Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
 - 50 UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
 - 51 Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande PT1749-016, Lisboa, Portugal
 - 52 Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande 1749-016, Lisboa, Portugal
 - 53 Department of Astronomy, University of Geneva, ch. d’Ecogia 16, 1290 Versoix, Switzerland
 - 54 INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
 - 55 INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
 - 56 Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
 - 57 Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
 - 58 Institut de Ciències de l’Espai (IEEC-CSIC), Campus UAB, Carrer de Can Magrans, s/n Cerdanyola del Vallés, 08193 Barcelona, Spain
 - 59 FRACTAL S.L.N.E., calle Tulipán 2, Portal 13 1A, 28231 Las Rozas de Madrid, Spain
 - 60 INAF-Osservatorio Astronomico di Padova, Via dell’Osservatorio 5, 35122 Padova, Italy
 - 61 Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
 - 62 Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, 0315 Oslo, Norway
 - 63 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

- 64 Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
- 65 Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- 66 Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
- 67 Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- 68 NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
- 69 Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- 70 Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Helsinki, Finland
- 71 Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- 72 Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
- 73 Department of Physics, P.O. Box 64, 00014 University of Helsinki, Helsinki, Finland
- 74 Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
- 75 NOVA optical infrared instrumentation group at ASTRON, Oude Hoogeveensedijk 4, 7991PD Dwingeloo, The Netherlands
- 76 Centre de Calcul de l'IN2P3/CNRS, 21 avenue Pierre de Coubertin, 69627 Villeurbanne Cedex, France
- 77 Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- 78 INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- 79 Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- 80 INFN-Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
- 81 Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- 82 Department of Physics, Institute for Computational Cosmology, Durham University, South Road, DH1 3LE, UK
- 83 Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France
- 84 European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- 85 Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- 86 Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, DK-8000 Aarhus C, Denmark
- 87 Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- 88 Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov 077125, Romania
- 89 Departamento de Astrofísica, Universidad de La Laguna, 38206 La Laguna, Tenerife, Spain
- 90 Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- 91 Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- 92 Université St Joseph; Faculty of Sciences, Beirut, Lebanon
- 93 Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada, 2008 Santiago, Chile
- 94 Satlantis, University Science Park, Sede Bld, 48940 Leioa-Bilbao, Spain
- 95 Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- 96 Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- 97 Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- 98 Dipartimento di Fisica, Università degli studi di Genova, and INFN-Sezione di Genova, via Dodecaneso 33, 16146 Genova, Italy
- 99 Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- 100 INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
- 101 Astronomical Observatory of the Autonomous Region of the Aosta Valley (OAVdA), Loc. Lignan 39, I-11020 Nus (Aosta Valley), Italy
- 102 Junia, EPA department, 41 Bd Vauban, 59800 Lille, France
- 103 ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
- 104 Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- 105 CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland OH 44106, USA
- 106 Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- 107 Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
- 108 Departamento de Física Fundamental, Universidad de Salamanca, Plaza de la Merced s/n., 37008 Salamanca, Spain
- 109 Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- 110 Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo, Kashiwa, Chiba 277-8583, Japan
- 111 Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- 112 Minnesota Institute for Astrophysics, University of Minnesota, 116 Church St SE, Minneapolis, MN 55455, USA
- 113 Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- 114 Department of Physics & Astronomy, University of California Irvine, Irvine, CA 92697, USA
- 115 Department of Astronomy & Physics and Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia B3H 3C3, Canada
- 116 Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- 117 Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- 118 CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bât. 709, 91191 Gif-sur-Yvette, France
- 119 Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- 120 Department of Astronomy, University of Florida, Bryant Space Science Center, Gainesville, FL 32611, USA
- 121 Department of Computer Science, Aalto University, PO Box 15400, Espoo FI-00 076, Finland
- 122 Instituto de Astrofísica de Canarias, c/ Via Lactea s/n, La Laguna E-38200, Spain. Departamento de Astrofísica de la Universidad de La Laguna, Avda. Francisco Sanchez, La Laguna E-38200, Spain
- 123 Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- 124 DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
- 125 Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000 Grenoble, France
- 126 Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Turku, Finland
- 127 Serco for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafraña del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
- 128 ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
- 129 Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia

- ¹³⁰ School of Physics and Astronomy, Queen Mary University of London, Mile End Road, London E1 4NS, UK
- ¹³¹ ICTP South American Institute for Fundamental Research, Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil
- ¹³² IRFU, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette Cedex, France
- ¹³³ Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden
- ¹³⁴ Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK
- ¹³⁵ INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
- ¹³⁶ Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy
- ¹³⁷ Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
- ¹³⁸ HE Space for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
- ¹³⁹ Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
- ¹⁴⁰ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- ¹⁴¹ Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- ¹⁴² Theoretical astrophysics, Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden
- ¹⁴³ School of Physics & Astronomy, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
- ¹⁴⁴ Department of Physics, Royal Holloway, University of London, TW20 0EX, UK
- ¹⁴⁵ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
- ¹⁴⁶ Department of Physics and Astronomy, University of California, Davis, CA 95616, USA
- ¹⁴⁷ Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
- ¹⁴⁸ Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
- ¹⁴⁹ Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
- ¹⁵⁰ Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010 New York, NY, USA

Appendix A: Morphological classification of the proto-cluster profile

In Section 5, we have explored the variation of proto-cluster profiles in the simulation lightcones. Here, we present figures for the mean profiles for each of the five categories in Figs. A.1 and A.2. The figures also show the standard deviations of the profiles in each category. Fig. A.3 shows examples of proto-clusters in categories 3 to 5. All these proto-clusters show large substructures or even multi-modality.

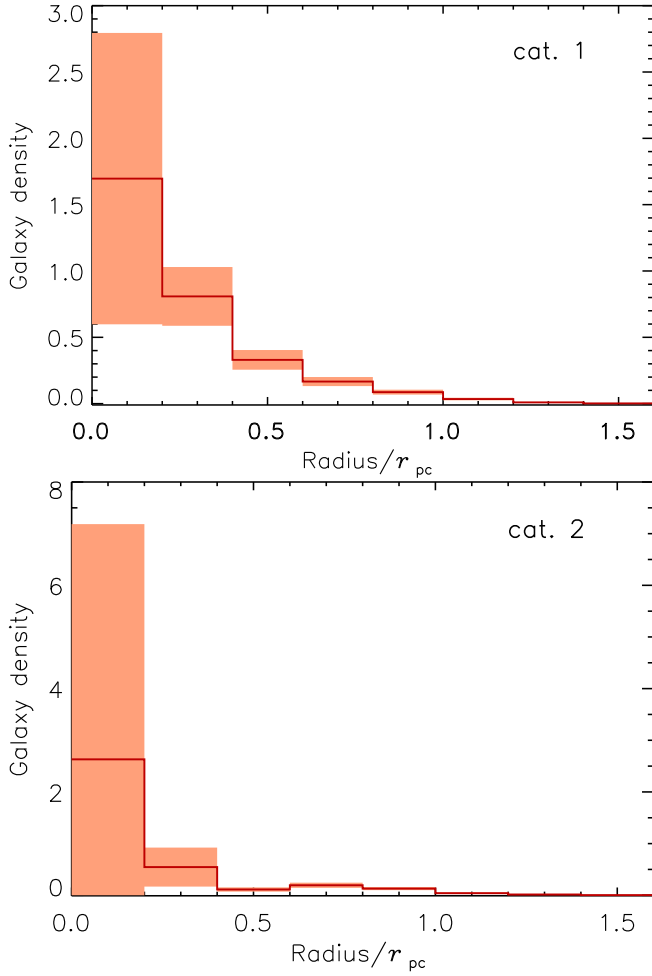


Fig. A.1. Mean profiles of the proto-clusters for two of the five structural categories defined in the text, where the categories are labelled in the panels. The shaded regions in the histograms indicate the rms variance of the density.

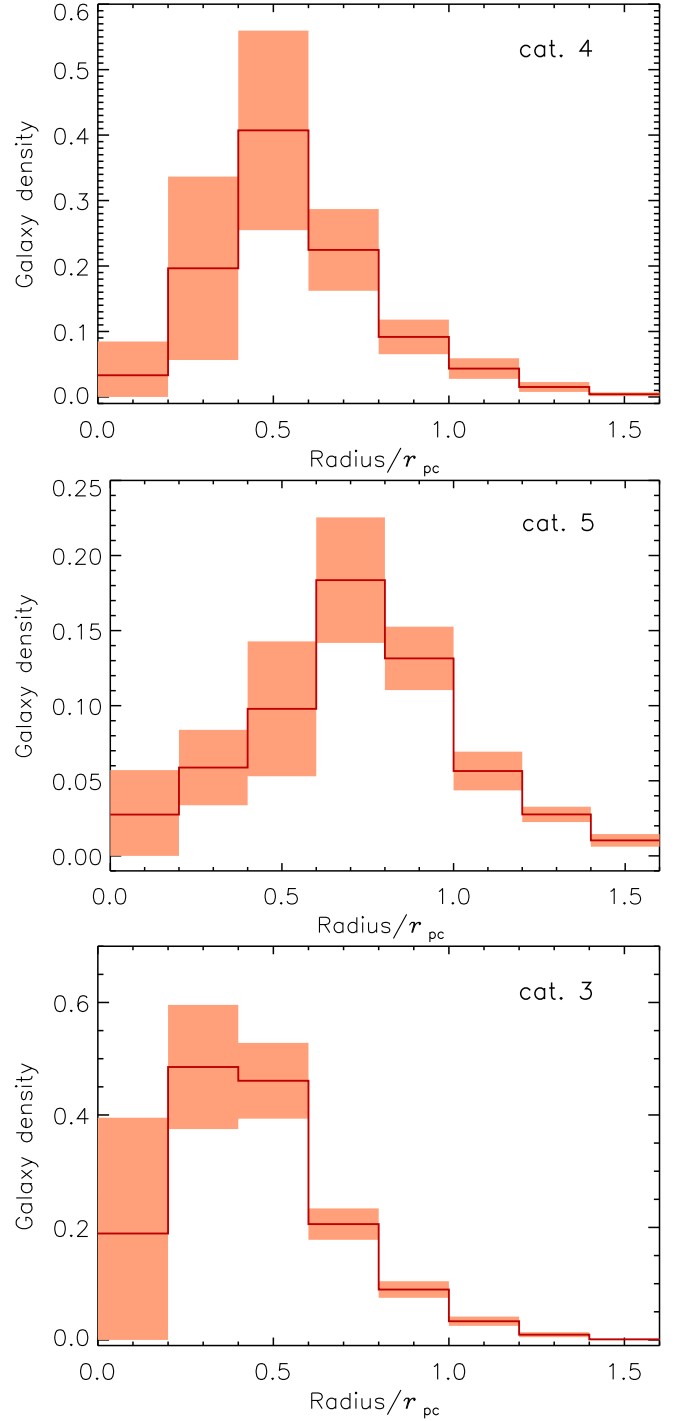


Fig. A.2. Mean profiles of the proto-clusters for the remaining three of the five structural categories defined in the text, where the categories are labelled in the panels.

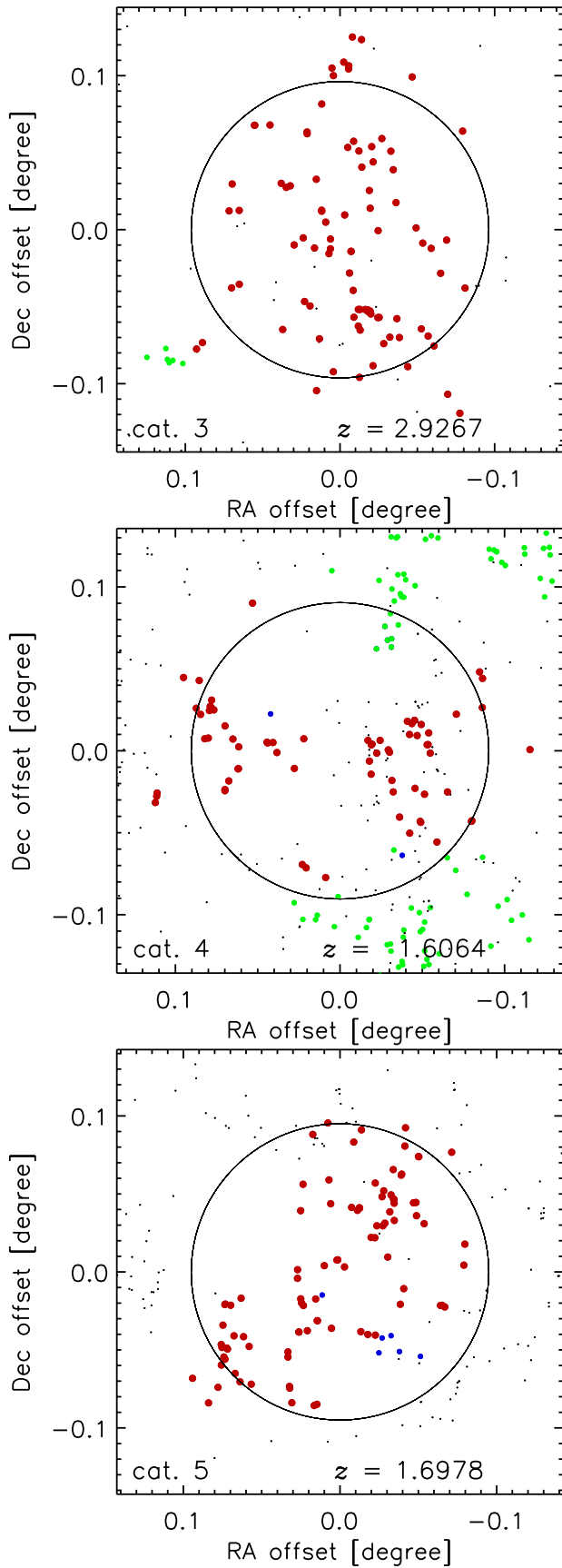


Fig. A.3. Examples of three proto-clusters for the categories 3 to 5. The category and the redshift of the proto-clusters is given in the legend of the Figure.