# Preparation for the analysis and interpretation of Euclid galaxy clusters

Alejandro Jimenez Munoz

Université Grenoble Alpes

**THÈSE**

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

Spécialité : Physique Subatomique et Astroparticules

Arrêté ministériel : 25 mai 2016


Présentée par

**ALEJANDRO JIMÉNEZ MUÑOZ**

Thèse dirigée par **Juan-Francisco MACÍAS-PÉREZ**

préparée au sein du **Laboratoire de Physique Subatomique et Cosmologie**
dans **l'École Doctorale de Physique**


# Préparation à l'analyse et interprétation des amas de galaxies avec Euclid

# Preparation for the analysis and interpretation of Euclid galaxy clusters

Thèse soutenue publiquement le **4 Novembre 2021**, devant le jury composé de :

**Monsieur Ken GANGA**
Directeur de Recherche, APC, Rapporteur
**Madame Sophie HENROT-VERSILLÉ**
Directrice de Recherche, IJCLab-Orsay, Rapporteuse
**Monsieur Rémi ADAM**
Chargé de Recherche, LLR, Examinateur
**Madame Hélène COURTOIS**
Professeure, IPNL, Examinatrice
**Monsieur Laurent DEROME**
Professeur, LPCS, Président
**Monsieur Juan-Francisco MACÍAS-PÉREZ**
Directeur de Recherche, LPCS, Directeur de thèse

# Table of Contents

# Acknowledgements

Starting these acknowledgements is not easy, but I hope not to forget anyone (sorry in advance, just in case). By the way, for those of you who can't speak Spanish: sorry not sorry.

First, I would like to thank the members of the jury: Ken Ganga, Sophie Henrot-Versillé, Rémi Adam, Hélène Courtois and Laurent Derome. Thanks for all the kind words, the advises, corrections and suggestions.

Ahora te quiero agradecer a ti, Juan. Ha sido increíble compartir este viaje contigo y sentir un pedacito de Andalucía cada día, aun estando tan lejos de casa. No ha sido tan fácil llegar hasta aquí, y tú lo sabes bien, pero siempre has sabido escucharme y entenderme, has sabido ayudarme y has estado dispuesto a adaptarte a cada inconveniente, e incluso has conseguido hacer que me lo pase bien durante la tesis (y me has aguantado mucho, mucho, mucho... Ni mis padres me soportan tanto). Gracias a ti, no abandoné la tesis, cuando pensé que era la única opción. Desde que empezamos a trabajar juntos en noviembre de 2018, hasta hoy, hay una cosa que no ha cambiado: mi gratitud hacia ti. Solo repetiría este viaje, si lo volviese a compartir contigo; y para ser sinceros, nos lo hemos pasado muy bien trabajando juntos, incluso con todo el drama que conlleva una tesis. Muchas gracias, Juan; no podría haber tenido un mejor director de tesis.

Por otra parte, quiero mencionar a quien me hizo irme a Grenoble para quedarme, enseñándome tanto sobre física, como, más importante aún, tanto sobre la vida. Mar Bastero Gil, aun conociéndote de tan poco tiempo, me has dejado huella. Me encanta cada vez que nos vemos y nos ponemos a hablar durante horas. Me encanta el hecho de que después de cada conversación siempre acabo aprendiendo algo. Porque tu enseñas siempre, incluso sin quererlo. Adoro eso de ti. Bueno, y porque a dar turra no nos gana nadie (había que decirlo). Muchas gracias, Mar.

I wanna thank also to Gustavo Yepes, Marco de Petris and, of course, Juan to the great opportunity I have been given to work on the Three Hundred Project and to stay at Rome for several months (grazie mille Marco). I just felt so in love with Italy. It has been a pleasure to work with you all, and to learn from you. Bueno Gustavo a ti también te lo digo en español: ¡Muchas gracias!

Os toca a vosotros, las personas más importantes de mi vida: mamá y papá. El problema con vosotros no es daros las gracias, es cuándo parar de dárosla. Con tantas necesidades como ha habido en casa, jamás me ha faltado nada, incluso cuando os teníais que sacrificar vosotros, para dárnoslo a nosotros. Sois las personas más generosas que conozco, y no podría estar aquí sin vuestro apoyo incondicional. Sé que estáis orgullosos de mí, de cada sueño que he cumplido con tanto esfuerzo, pero yo estoy aún más orgulloso de vosotros, de ser vuestro hijo. Gracias, gracias y gracias. Jamás me cansaré de decirlo. Os quiero, mamá y papá. Por supuesto, tengo que agradecer a todo el resto de mi familia: a mis hermanos, mis cuñadas y a mi pequeño Raúl; mis mellis y a mis tías; y sobretodo a mi tita Mari Loli, que me ha alimentado durante mi año de máster en Granada, que me dió tanta compañía aquellos domingos cuando nos tomábamos un vargas y hablábamos sobre tantas cosas. Gracias, tita, por ayudarme a sobrevivir. Y quien también me ayuda a sobrevivir es mi madrina, mi Toñi. Gracias, tita, por tener siempre palabras de cariño y apoyo hacia mí. Al final del día, resulta ser que estoy muy bien acompañado, aunque casi siempre sea en la distancia. Os quiero, os quiero tanto a todos.

Hermana mía, sí, tú, es tu turno. ¿Cuánto significas para mí? Demasiado como para poder expresarlo con palabras, pero lo voy a intentar. Eres mi hermana, mi mejor amigo, mi psicólogo (encima de gratis), mi confidente, mi mayor apoyo... En definitiva, lo eres todo para mí. Eres quien me rescata cuando necesito ayuda, aun cuando ni yo sé que la necesito. Has estado a mi lado desde que empezamos la carrera en 2013. No recuerdo estudiar ningún examen de la carrera sin ti, ni siquiera ir a clase y no tenerte cerca de mí. Cada momento de felicidad que he tenido en mi vida desde 2013 te incluye a ti. Hemos compartido cada lágrima, pero también cada risa. No recuerdo pasar ni un día sin reírnos ni sin hacer gilipolleces, ya sea en Córdoba, San Sebastián, Roma, Madrid... En cualquier sitio me siento como en casa, si estoy contigo. Fue precisamente en San Sebastián, cuando descubrimos que yo iba a hacer un doctorado, que tenía que volver a Francia y que empezaba este viaje tan bonito; pero lo mejor de todo fue poder darte un abrazo y tenerte a mi lado cuando nos enteramos, compartir ese momento que tan feliz me hizo, y que tanto miedo me daba. A día de hoy puedo decir que te necesito en mi vida, y lo digo orgulloso. Eres imprescindible para mí; imprescindible por todo lo que nos une, que no es solo todo lo que he mencionado anteriormente, sino que por si fuera poco pertenecemos al mismo maravilloso colectivo: al LGTBIQ+. Cada día me animas a aceptarme y quererme tal y como soy, a expresarme y sentirme libre y a que viva mi vida independiente del juicio ajeno. Soy feliz siendo gay, y aún más sabiendo que puedo compartir eso contigo, con mi hermana. Solo contigo puedo cantar *Crazy in Love* sin camiseta, cogidas de la mano y haciendo todo un desfile por una discoteca en Milán; o gritar a pleno pulmón en la Gran Vía de Madrid: '*Y TODOS ME MIRAN, ME MIRAN, ME MIRAN, PORQUE SÉ QUE SOY LINDA, PORQUE TODOS ME ADMIRAN*'. Y así podría seguir durante horas, porque tenemos demasiados momentos juntos,

aunque queden millones más por venir. Tengo la mayor suerte del mundo por tenerte. Pero como esto va de agradecer, te quiero dar las gracias por ser el mejor de los amigos, un tesoro con un valor incalculable, el que me nutre intelectualmente, gracias no solo a tu inteligencia, sino a todos los debates que compartimos, con el que puedo hablar de introspección, de evolución personal, de deconstrucción (en lo que, por cierto, eres mi referente), de libertad, y de un infinito etcétera. Ojalá sea como tú cuando sea mayor. Gracias infinitas, lo eres todo, y siempre lo serás. Te quiero.

PD: Seis nuggets.

PPD: No me puedo olvidar de ti, mamá adoptiva. Mañanas, tardes y noches de turra. Turra en la que tanto a Migui como a mi, nos ayudabas a gestionar (o intentarlo) todo lo que significa hacer una tesis, que no es poco. Y por aguantarme tantos años de ocupa en tu casa. ¿He molestado? SI. ¿Sigo molestando? TAMBIÉN. ¿Seguiré molestando? Eso no lo dudes. ¿Te quiero? Mucho mucho mucho. Gracias a ti también Mercho, aunque ya no estés enamorada de mi, yo te quiero mucho.

Let's switch to English (I know you wanted to, my people), and let's go to Grenoble. This thesis has been an amazing journey with so many personal rewards. Without question, you are the biggest one, Rolita. Our friendship started in the most random, and yet, the best way to describe us: crying with a beer, at night, and talking about drama... This is just us. From that day on, we were inseparable. My biggest luck was to be with you every single day, both at lab, and outside. Thanks to you I discovered the true meaning of regret: doing break after break and realizing I did nothing on the full day. Solution to this regret? Meeting after the lab to go to the cinema. Yes, again, this is just us. The last year of my thesis, I was missing so much your visits to my office, and your emotional support every day. But the conclusion from all of this was: OK... I need Rola, I love Rola, I want Rola on my life F-O-R-E-V-E-R. Spoiler: This continues to be the case. Our friendship could be defined by three questions that I've made myself during these years, with just one answer. The first question, which came to my mind not so long after meeting you, was: What would I do without Rola in the lab? Soon, It came the second question: What would I do without Rola in Grenoble? The last one, which still remains right now, is: What would I do without Rola in my life? Curiously the answer to the three questions is the same: I wouldn't do anything without you. Thank you, for being there for me, for being such an amazing friend, for sharing with incredible moments that are irreplaceable. I love you, 7ayete.

Of course, I couldn't speak of Grenoble, without speaking about you, bitch. The fucking Salah, which name I won't be able to pronounce, ever. My lovely little baby. I think I managed to finish the thesis thanks to our Thursday-night-sushi moment. Or our Friday-night-Vietnamese

moment. Or our whatever day-night-restaurant combination moment. I think the workers of these places know everything about our life, because we were serving drama as nobody is able to do (and yes, I don't know if I'm making up this expression, but you know, perfectly, what I mean). How many calls we've done to each other just saying: I'm picking you up downstairs, let's go to take a walk, I need it ? Infinite is not even closer. You, as every other tormented PhD student, like me, needed support, and we found, together with Rolita, the best trio of friends. I miss so much, when we were together the three of us, surrounded by people, and just with a look we could, perfectly, understand each other. I love how, even waiting on a queue for getting the first shot of vaccine (not even with rendez-vous), we are able to have so much fun. Salah, I remember how, in the worst period of the PhD, when I was writing this wonderful manuscript, you were coming with me to work, to give me support, and help me not to give up. And what I still love, while I'm writing this, it is that I'm right now planning trips with both of you, Rola and Salah, and I know this friendship will last forever. I received the best gift when I decided to go to Grenoble: both of you. When I'm feeling sad or alone, I always read the cards you gave me as a present for the PhD, and I want to give you back a card. What I love the most about you? That both of you are still in my life. Thank you so much. I love you so so much.

Debería de ir acabando, pero es que resulta que estoy muy bien acompañado en la vida, y tengo tanto por lo que agradecer. Así que aun queda un poco. (Important translation: I should be finishing up, but I happen to be in very good company in life, and I have so much to be thankful for. So there's still a little bit left to say).

Antes he hablado de Granada; también de Italia. Y no puedo no pensar en ti, Ruth. Nos conocimos por casualidad (no diré cómo, para mantener nuestra buena reputación) y desde entonces no hay quien te eche de mi vida. Resulta que cuando me mudo a un sitio, la primera persona que busca un vuelo para venir a verme (y acaba viniendo) eres tú; pero es que resulta que yo he cogido la misma buena costumbre contigo. Ya se me hace raro no organizar una Navidad contigo, donde casi siempre acabamos solos, comiendo bonito, chorizo, jamón, tortilla... El kit básico español, vaya. Además, siempre nos solemos hacer un regalo (sí, tú eres mejor regalando que yo, lo sé) y siempre tienes la manía de engañarme, para hacerme elegir mi propio regalo sin yo saberlo. Parece ser que no soy demasiado listo después de todo. ¿Qué nos caracteriza a nosotros como amigos? Bueno, pues nos encanta ver realities en Netflix, y nos gusta cotillear a tope con ellos. Nos encanta viajar, y nos encanta visitarnos cada dos por tres. También nos encanta visitar una página web maravillosa, en la que, por tu culpa, me suelo arruinar cada vez que entro (www.platanomelon.com). Nos encanta estar hasta las tantas, hablando y hablando y hablando... Si es que al final pensamos igual, así que es muy fácil hablar tantas horas. Nos encanta emborracharnos juntos, aunque yo creo que nos pasamos... Bueno, los dos nos hemos

caído de una bici en Italia, y no una bici cualquiera no, una Mobike. Eso une mucho. Creo que mejor paro ya, ¿no? Porque podría seguir así todo el día. Muchas gracias amiga, por aguantarme tanto, pero siempre con una sonrisa. Que eso, que te quiero, que vayas buscando vuelos para venir a verme.

Let's finish ! Finally, I know. I wanna thank to the amazing people of the LPSC. Starting with Florian, because you had to share with me an office during 3 years, I'm sorry ! But we had so much fun, and you helped me a lot during my thesis, so a big thank for you. Also the rest of the COSMO ML group and DARK group: Fred, Laurence, Céline, Laurent, David. You welcomed me in the best way possible, and you always had patience when I was trying to speak french (so sorry), merci beaucoup ! Of course the PhD and Post-doc students of the lab (those who have already left, and those who still remain): Flora, Killian, Emmanuel, Calum, Carolina, Miren, Alessandro... Por cierto Miren, increible el póster que preparaste para mi defensa de tesis. Espero que Juan no te moleste mucho. Mucha suerte, hay que tener paciencia con él (es broma, Juan. Te queremos mucho). Quiero darles las gracias también a Dilia y Carolina, mi pequeño refugio en Francia, donde podía hablar español y sentirme en casa. Después de todo, parece ser que los andaluces y los colombianos no somos tan diferentes. I wanna thank also, Raphael, for being my tea supplier at the lab and also for hosting so many dinners (in which sometimes we had to fight with Deliveroo). I don't want to forget to mention everybody I met outside the lab: John, Marta, Giovanna, Rohan... Thanks to all of you! A mis niñas de Córdoba, Gemma y Laura, os doy las gracias por tantos años de amistad. También te doy las gracias a ti Javi, amiga. El confinamiento nos hizo descubrirnos, por segunda vez, y fue maravilloso. Thanks to everyone. What an amazing journey!

# INTRODUCTION

**You never completely have your rights, one person, until you all have your rights.**

**Marsha P. Johnson**

Cosmology is the science that studies the origin and evolution of the Universe. This field has grown up exponentially in the last decades thanks to the development of numerical simulations and observational surveys, both covering a significant region of the sky. The current cosmological model has been validated numerous times thanks to these observations, with galaxy clusters being a key pillar both for the development of the model and to constraint it. One of the major advantages of galaxy clusters is that they have formed late in time, and they are the most massive bounded structures in the Universe. The other great advantage is that they can be observed and studied in a wide range of the electromagnetic spectrum. For example, the ionized gas, that makes up 12% of their mass, gives us information at millimetre and X-Ray wavelengths, as demonstrated by the *Planck*, XMM and Chandra satellites, which have given an unprecedented view of clusters of galaxies and cluster cosmology. The most relevant current and future surveys in this domain will be eROSITA (X-Ray) and CMB-S4. In the visible and infrared part of the electromagnetic spectrum, galaxy clusters can be detected from the light emitted by their stars and galaxies, which only makes up 3% of the total mass of the clusters. Historically, clusters of galaxies were first detected at these wavelengths. Current surveys like SDSS and KiDS are producing already great results in cluster cosmology. Soon the Vera Rubin Observatory, and the *Euclid* satellite, will allow us to explore a wide range of the Universe with great precision, by means of its stellar and galactic component. The surveys will detect hundreds of thousands of clusters, leading to a precise cosmology with cluster member counts. One of the major challenges will be to define the selection function for these surveys. In order to define this function, it is necessary to use numerical simulations that reproduce the technical and scientific requirements of the specific survey as well as the properties of the clusters. The first numerical simulations used in the context of cosmology were developed in the 60s and included only non-baryonic physics, called dark matter only simulations. It was not until the 90s when, thanks to the advance of computational power, it was possible to include baryonic physics in numerical simulations, called hydrodynamical simulations. Recently, The Three Hundred collaboration has developed a cluster catalogue constructed with hydrodynamical simulations with high statistics

and resolution, taking into account the difficulty of generating this type of simulation.

Within this context, the *Euclid* satellite will make pioneer observations in the field of cosmology, thanks to the large region of the sky it will observe, the incredible precision of its measurements, and the acquisition of data in the infrared for these regions, which is impossible or very difficult to access from Earth. Prior to its launch, it is necessary to properly characterize its physical components, as well as to prepare the cosmological analysis. The work performed in my thesis focuses on cosmology with galaxy clusters and the characterization of the infrared detectors of the NISP instrument. My thesis is organized in three main parts. Part I introduces the general context:

— Chapter 1 presents the concordance cosmological model focusing in the process of formation of large scale structures.

— Chapter 2 introduces galaxy clusters and their use in cosmology. We describe the link between the number of clusters as a function of the mass and redshift with the cosmological parameters, with a brief state of the art of the latest cosmological results. We present several galaxy cluster observables and mass proxies, paying special attention to the optical/infrared domain. Finally, I explain, briefly, how to detect a cluster and compute the survey selection function.

In Part II we describe the *Euclid* satellite from its components to the characterization of its infrared detectors.

— Chapter 3 presents the *Euclid* mission. We describe the technical characteristics of the telescope and its instruments: the visible instrument (VIS) and the near infrared instrument (NISP) and its scientific goals.

— Chapter 4 describes the implication of correlated readout noise for flux measurement with the *Euclid* NISP instrument.

Part III describes the complex process to determine the selection function of a cluster catalogue through a cluster injection method.

— Chapter 5 characterizes two main observational properties using the *Euclid* Mock catalogue: the galaxy density radial distribution and the luminosity function; by fitting them with two main analytical models, a Navarro-Frenk-White (NFW) distribution and a Schechter function.

— Chapter 6 presents the construction of a synthetic cluster catalogue based on the results of Chapter 5.

— Chapter 7 describes a cluster injection method to detect clusters. We describe the PZWAV cluster finder and we studied its performance on several cluster catalogues by computing the completeness and purity. Finally, we present an attempt to estimate the *Euclid* cluster catalogue selection function.

— Chapter 8 presents the Three Hundred Project, a 324 cluster sample simulated with full-physics hydrodynamical re-simulations, which could be used in the context of *Euclid*. We recover the same galaxy properties discussed above in a more realistic way. We discuss the impact of resolution effects in these properties as well as the impact of baryonic physics in the structure formation process.

Part I

# General Context

Chapter 1

# Theoretical Framework

In this chapter we present the theoretical framework in which this thesis is based, starting from the standard cosmological model to the description of some cosmological probes and in particular those related to galaxy clusters.

## 1.1 Standard cosmological model

The theoretical framework of the concordance cosmological model describes gravity interaction through Einstein's General Relativity [1]. Considering an isotropic and homogeneous Universe, Einstein's equations can be solved leading to the equations of dynamics [1] of Friedmann-Lemaitre-Robertson-Walker (FLRW) [2, 3]:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1.1}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(3p + \rho), \tag{1.2}$$

where $G$ is the Newton's gravitation constant, $k$ is the space-time curvature, $\rho$ and $p$ are the density content and pressure present in the Universe, $a$ is the scale factor and $\dot{a}$ its time derivative. Assuming three components: radiation, matter and a cosmological constant (dark energy) the total energy density of the Universe can be written as:

$$\rho = \rho_m + \rho_r + \rho_\Lambda, \tag{1.3}$$

$\rho_m$ being the matter density, $\rho_r$ the radiation and $\rho_\Lambda$ the dark energy density related to the cosmological constant $\Lambda$, defined as $\rho_\Lambda = \frac{\Lambda}{8\pi G}$. The expansion rate of the Universe is given by the Hubble constant, $H$, defined as the logarithmic derivative of the scale factor, $H = \dot{a}/a$. Assuming that the Universe components behave like perfect fluids, which state equation $w = p/\rho$, with $\rho$ and $p$ the density and pressure, we can write

$$\dot{\rho} + 3H(1 + w)\rho = 0. \tag{1.4}$$

1. In this chapter the convention of $c = 1$ is used.

Solving equation 1.4 for each energy density component, it is possible to obtain the dependency of the energy density with the scale factor (and thus, with time):

$$\rho(a) \propto \begin{cases} a^{-3} & \text{for } w = 0\text{: matter (pressureless)} \\ a^{-4} & \text{for } w = 1/3\text{: radiation} \\ a^{-3(1+w_\Lambda)} & \text{for } -1 < w_\Lambda < -1/3\text{: dark energy.} \end{cases} \tag{1.5}$$

For a flat Universe, where $k = 0$, and using the Hubble parameter dependency with the scale factor, and equation 1.1, we can define the critical density $\rho_c$ as:

$$\rho_c = \frac{3H^2}{8\pi G}. \tag{1.6}$$

Thus we can define the ratio of the density to the critical density, also called normalized density, $\Omega_i$, as:

$$\Omega_i = \frac{\rho_i}{\rho_{crit}}, \tag{1.7}$$

where $\rho_i$ represent different density components, e.g., $i = m$ for matter density or $i = r$ for radiation density. In terms of this variable, the Friedmann equation 1.1 can be rewrite as:

$$\Omega_{tot} - 1 = \frac{k}{a^2 H^2}, \tag{1.8}$$

where $\Omega_{tot}$ is the total normalized matter-energy content of the Universe and $\Omega_k = \frac{k}{a^2H^2}$ the normalized space curvature density. For different values of the space curvature $k = -1, 0, 1$ or, equivalently, $\Omega_{tot} > 1$, $\Omega_{tot} = 1$ and $\Omega_{tot} < 1$, the Universe is closed, flat or open, respectively. Equation 1.1 can be rewrite in terms of the scale factor, $a$, the Hubble parameter, $H$, and the normalized density $\Omega$ by:

$$H(a) = H_0 \sqrt{\Omega_{\Lambda,0} a^{-3(1+w_{DE})} + \Omega_{m,0} a^{-3} + \Omega_{r,0} a^{-4} + \Omega_{k,0} a^{-2}}, \tag{1.9}$$

where the subscript zero represents the value of any parameter at the present time. Thus $H_0$ is the expansion rate and $\Omega_{k,0}$ is the curvature density term today. The equation for the expansion rate as a function of the scale factor will be useful due to the fact of the relationship between the geometrical information within the scale factor and how to measure distances in the Universe. We expect the Universe to expand and to be dominated by radiation at its early stage, then by matter. Current cosmological constraints [4] also show the Universe dynamics is now dominated by dark energy, compatible with a cosmological constant.

## 1.2 Distance Measurements

One major issue in observational cosmology is how to measure distances. It is useful to define the redshift, $z$, as the ratio of the light wavelength of a source at the time the light is emitted, $\lambda_{emit}$, and at the time it is observed, $\lambda_{obs}$:

$$1 + z = \frac{\lambda_{obs}}{\lambda} = \frac{a_0}{a}, \tag{1.10}$$

where by definition of the scale factor, $a = 0$ is the origin of the Universe that corresponds to high redshifts, up to nowadays, $a = 1$ corresponding to $z = 0$. Now equation 1.9 can be written in terms of $z$ as:

$$H(z) = H_0\sqrt{E(z)}, \tag{1.11}$$

where $E(z)$ is defined by:

$$E(z) = \sqrt{\Omega_{\Lambda,0}(1+z)^{3(1+w_{DE})} + \Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4 + \Omega_{k,0}(1+z)^2}. \tag{1.12}$$

A way to define the distance between an observer at $z = 0$ and a cosmological object at $z$ is using the radial comoving distance, $D_C$, defined by:

$$D_C = \int dt \frac{1}{a(t)} = \int_0^z dz' \frac{1}{H_0 E(z')}. \tag{1.13}$$

The term "comoving" refers to variables that are invariant with respect to the expansion of the Universe. To measure the "physical" or proper distance it is necessary to take into account the scale factor. Thus, the radial proper distance is defined as $d(t) = aD_C$. The radial comoving distance is measured in the line-of-sight, but to measure two cosmological objects at the same redshift, $z$, it is necessary to define the transverse comoving distance, $D_M$, which depends on the space curvature, and it is defined as follows:

$$D_M = \begin{cases} L\sinh\left(D_C/L\right) & \text{for } \Omega_k > 0 \\ D_C & \text{for } \Omega_k = 0 \\ L\sin\left(D_C/L\right) & \text{for } \Omega_k < 0 \end{cases} \tag{1.14}$$

where $1/L = H_0\sqrt{|\Omega_k|}$. Again, the transverse proper distance, which accounts for the Universe's dynamics is defined as $d_M = aD_M$. The latter is the physical size of a gravitationally bound object. Now we can define the angular diameter distance of an object at redshift z, as the ratio

between the physical size of an object and its observed angular size:

$$D_A = \frac{D_M}{1+z}. \tag{1.15}$$

Finally, we define also the luminosity distance that relates the intrinsic luminosity of an object, $L$ and its flux $F$ as $D_L = \sqrt{\frac{L}{4\pi F}}$, and in terms of the redshift of the object is given by:

$$D_L = (1+z)D_M = (1+z)^2 D_A. \tag{1.16}$$

## 1.3 Thermal History of the Universe

The concordance cosmological model is based on the "Big Bang" theory developed by George Lemaître [5, 6, 7] and Alexander Friedmann [2] independently. The Universe is considered isotropic and homogeneous and a dynamical object where the Einstein's General Relativity applies, as discussed before. It expands with time and it was a hotter and denser at its beginning. The discovery by Edwin Hubble [8] that galaxies were moving away from us with increasing velocity when increasing distance to us, its electromagnetic spectrums are redshifted, constituting a fundamental observational proof for this cosmological theory. At the early universe the temperature of the Universe is high and particle interactions are important so that they are in the form of a hot plasma in equilibrium dominated by radiation [9]. When the Universe expands, it cools down to the point where, protons and neutrons combine to form hydrogen, then helium and after heavier nuclei, beginning a period of matter-dominated Universe, reaching the matter-radiation equality at $z \sim 3600$. This process is known as primordial nucleosynthesis and the measurement of the abundance of light elements [9] constitutes another fundamental observational proof for the Big Bang theory. At $z \sim 1100$ the interaction rate between photons and electrons is small with respect to the expansion rate of the Universe, and electrons and protons combined to form neutral atoms in what is called the recombination epoch. This leads to a neutral Universe and light propagates freely. This early radiation known as the Cosmic Microwave Background can be observed today. It was first discovered by Penzias and Wilson [10] in 1964. The COBE (Cosmic Background Explorer [11]) satellite measured the CMB to be a black body, isotropic and homogeneous as predicted by the Big Bang theory. The left part of Figure 1.1 shows the CMB electromagnetic spectrum measured by the instrument FIRAS of the COBE satellite. The black line corresponds to a black body of temperature of 2.725K that fits perfectly the data from COBE/FIRAS (see left part of Figuree 1.1). The CMB has been measured by other instruments with more precision like the satellites WMAP (Wilkinson Microwave Anisotropie Probe [12]) or Planck [13]. The right part of Figure 1.1 shows the CMB temperature anisotropies of rms 100 $\mu$K measured by Planck. This confirms that the Universe is homogeneous and isotropic at large angular scales. The CMB is another observational proof,

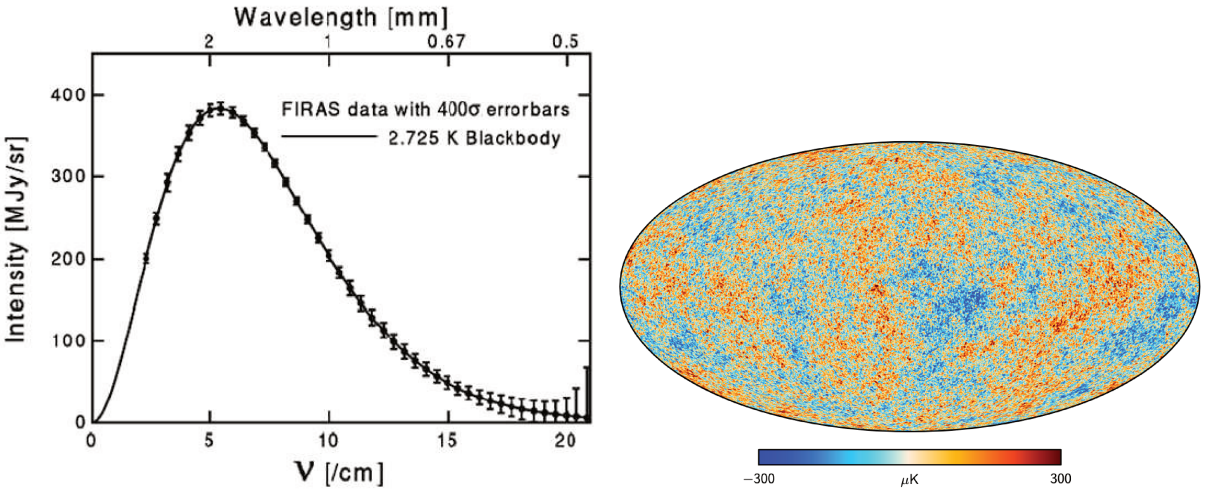


Figure 1.1 – Left: CMB Electromagnetic spectrum measured by the instrument FIRAS of the COBE satellite [14]. Right: CMB temperature anisotropies [13].

fundamental to accept the "Big Bang" model.

The fact that there are small anisotropies in the CMB, and at large scale the Universe is homogeneous and isotropic can not be explained by the Big Bang theory. A theoretical paradigm called inflation was proposed in the 80's [15, 16, 17] that postulates that at the early Universe there were small quantum fluctuations that, after a rapid expansion period, grew leading to the large-scale structures that are found nowadays such as galaxy clusters. The amplitude and spectral index of the power spectrum of these quantum fluctuations, $A_s$ and $n_s$, respectively, are part of the six independent cosmological parameters that defines the concordance cosmological model. In the latter, also called $\Lambda CDM$ for $\Lambda$ (dark energy) and CDM (Cold Dark Matter) postulates that the Universe consist of: dark energy (70%), cold dark matter (25 %), baryons (∼5%) and photons and neutrinos. From this we observe that 95% of the content of the Universe is unknown. The matter density of baryons and dark matter are two of the six independent cosmological parameters, $\Omega_b\ h^2$ and $\Omega_c\ h^2$, respectively. Dark matter has not been observed directly due to its of cold (relativistic) and collisionless (interacting just gravitationaly) nature, even though it constitutes the 25% [4] over the 30% of total matter (baryons and dark matter) present in the Universe.

Figure 1.2 shows a timeline of the story of the Universe following the cosmological concordance model, by ESA. After the CMB forms, matter dominates in a period that is called "dark ages". Dark matter collapses and forms halos and filaments where baryonic matter accumulates. In this dense spots the first stars appear at a redshift of about z = 20. After stars, later galaxies and finally galaxy clusters, that are the most massive gravitationally bound structures. The lat-

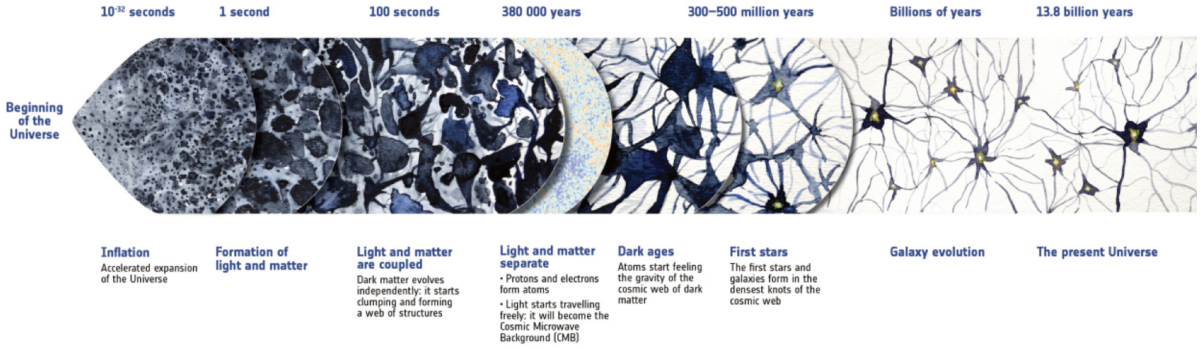


Table 1.2 – Timeline Universe. Credit: ESA, https://sci.esa.int/web/planck/-/55392-the-history-of-the-universe.

ter has been a fundamental object of study for the development of the concordance cosmological model. The measure of the velocities of galaxies in the Coma Cluster by Fritz Zwicky in the 20's-30's decade [18] lead to the proposition of a new type of invisible matter, known as Dark Matter, to explain the lack of mass necessary to explain the movement of the galaxies.

The matter dominated Universe ends at a redshift of about $z = 0.4$ where the Universe expansion accelerates [19]. The measure of distances from supernovas type SN1a, considered as standard candles in the late 90's [20, 21], demonstrated that the Universe is in an expansion phase dominated by Dark Energy [22].

## 1.4 Structure Formation

CMB measurements show that the Universe is homogeneous and isotropic to $10^{-5}$. These anisotropies in the CMB temperature map correspond to the initial density perturbations that lead to structure formation. These initial density fluctuations are seeded by quantum fluctuations at time of inflation.

The structures in the Universe are formed from these quantum fluctuations that generates primordial density fluctuations. We can define the density contrast at a position and time, $(\vec{r}, t)$ like:

$$\delta(\vec{r}, t) = \frac{\rho(\vec{r}, t) - \overline{\rho}(r, t)}{\overline{\rho}(t)}, \tag{1.17}$$

where $\rho$ and $\overline{\rho}$ are the matter density and its mean value at time, $t$. Initial fluctuations can be derived from a primordial power spectrum $P_{prim}(k) = A_s k^{n_s - 1}$ with $n_s = 0.966$ [4], which makes the power spectrum quasi-scale invariant. In the linear regime, the power spectrum evolves during the history of the Universe as function of the cosmology and can be expressed as

$$P_m(k, z) = P_{prim}(k) D^2(z) T^2(k), \tag{1.18}$$

$T(k)$ is the transfer function that describes the impact of the linear regime growth up to the recombination period $z \sim 1000$, and $D(z)$ is the growth factor and it is related to how the perturbations grows. These two quantities depend on the cosmological parameters. The r.m.s of the matter fluctuations at a mass scale, M, is related to the fluctuations power spectrum by:

$$\sigma^2(M, a) = \int \frac{d^3k}{(2\pi)^3} W(kR) P_m(k, a), \tag{1.19}$$

where $W(kR) = 3\left(\frac{\sin(kR)}{(kR)^3} - \frac{\cos(kR)}{(kR)^2}\right)$ is the window function for a sphere of radius, R. The largest gravitationally bound structures in the Universe are galaxy clusters whose size is the order of Mpc. For this reason, the amplitude of the matter power spectrum is normalised by the r.m.s at a distance R = 8 $h^{-1}$Mpc at redshift z = 0 (a = 1), and this normalization is known as $\sigma_8$. As we will see in the next section, the number of clusters as a function of the redshift and mass is sensitive to the cosmological parameters, thus cluster number counts constitutes a major cosmological probe. In the case where perturbations are not in the linear regime (small scales or later times) numerical simulations are fundamental to understand the non-linear perturbations regime. As an example, Figure 1.2 shows the dark matter density field on various scales. The zoom sequence shows a region four times smaller than the previous one, centered in a galaxy cluster. As seen in the simulation, galaxies and galaxy clusters are formed in the the connections of filaments in the cosmic web. The fluctuations that have grown up and produce gravitational

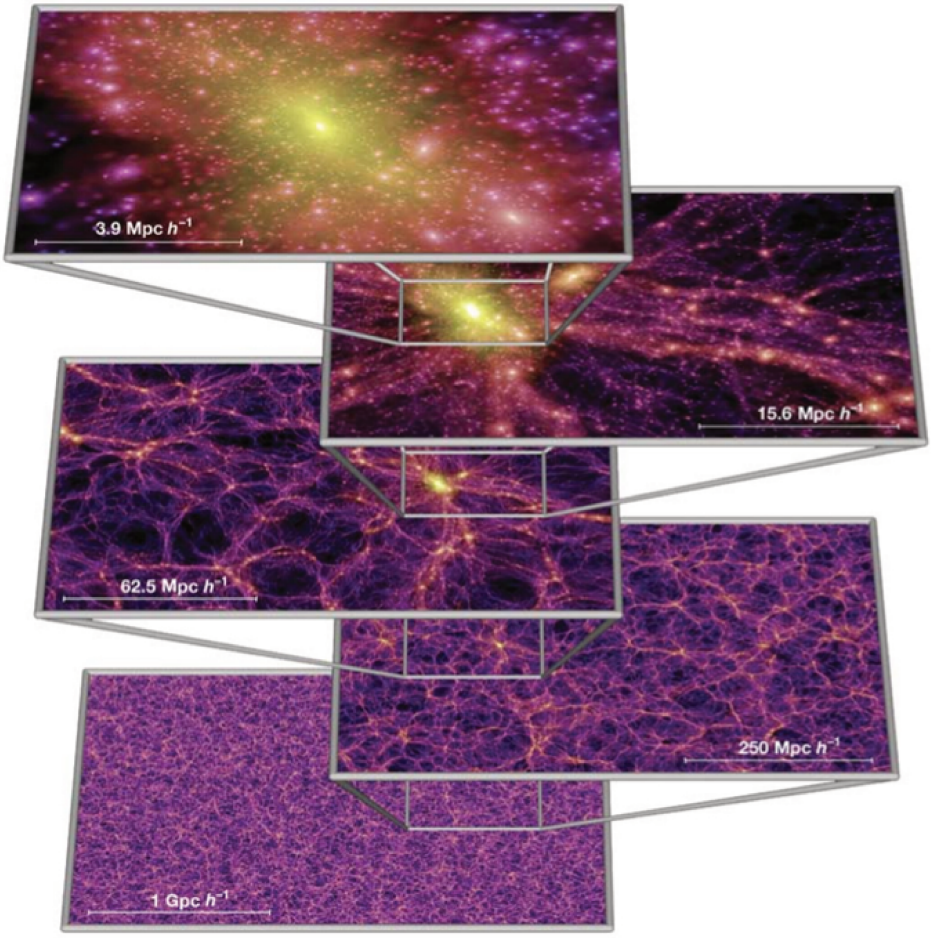


Figure 1.2 – Dark matter distribution from N-Body simulations. Credit: Millenium Simulation [23].

instabilities produces overdensities in the Universe, that collapse to form halos.

Following the analytical Press and Schechter formalism [24], halos collapse when the density contrast is $\delta \sim 1.68$ by gravitational effects and decoupling from expansion. Later on, the halo is virialised, meaning that the virial theorem, where the potential energy is twice the kinetic one, i.e., $E_p = -2E_k$, and the collapse equilibrium is reached. The number of halos as a function of mass and redshift can be obtained through the Halo Mass Function (HMF). The form of the HFM given by Tinker et al. [25] is

$$\frac{dN}{d\ln M} = f(\sigma)\frac{\overline{\rho}_m}{M}\frac{d\ln\sigma^{-1}}{dM}, \tag{1.20}$$

where $\overline{\rho}$ is the mean density of the Universe, $\sigma$ is computed from equation 1.19, and $f(\sigma)$ given

by

$$f(\sigma) = A\left[1 + \left(\frac{\sigma}{b}\right)^{-\alpha}\right] \exp\left(-\frac{c}{\sigma^2}\right), \tag{1.21}$$

where $(A, a, b, c)$ are parameters estimated from numerical simulations (e.g., [25]). Therefore, a cluster can be defined by its mass and redshift, and they are self-similar objects. We will see in the next chapter that the evolution of the number of halos as a function of the mass and redshift depend on the cosmological parameters, $\sigma_8$ and $\Omega_m$, thus galaxy clusters are a cosmological probe.

## 1.5 Numerical Simulations

The understanding of how a nearly uniform and isotropic Universe evolves to form stars, galaxies and large scale structures is a challenging problem in modern cosmology. The use of numerical simulations in the non-linear regime is fundamental, where the density fluctuations are comparable with the matter density, $\delta\rho \sim \rho$, and the gravitational evolution in a FLRW Universe can not be performed analytically. The development of numerical simulations from the 60's and their improvement and combination with observational measurements have been an achievement in modern cosmology. Now we are going to explain the different type of cosmological simulations and a bit of historical context for each one (for a more detailed review see [26, 27]).

### 1.5.1 N-body Simulations

N-body, dark-matter-only simulations consider generally gravity as the only interaction between particles, solving the Vlasov-Poisson equations. The first simulation of the evolution of N gravitating bodies (less than one hundred particles) was in the 60's [28, 29]. In the next decade the development of computers allowed researchers to increase the number of particles for studying the formation and evolution of cluster of galaxies [30, 31], Press and Schechter in the 70's studied the development of clustering in an expanding Universe by considering one thousand particles placed randomly in a sphere that is expanding, and compare the results of these simulations to their analytical interpretation [24]. The 70's decade ends with allowing to perform N-body simulation in expanding spheres containing up to 5000 particles [32].

Early in the 80's it became possible to compute an arbitrary power spectrum for the density fluctuations thanks to the application of the Zeldovich approximation [33] in two dimensions by Doroshkevich [34] and three dimensions by Klypin and Shandaring [35]. This is the standard method nowadays to set up the initial conditions in agreement with predictions from the linear theory. The previous works applied the Particle-Particle (PP) algorithm where the motion of particles is the direct summation of pairwise forces. However, they are not applicable for a large

number of particles, because the computation cost scales with $N(N-1)$ where $N$ is the number of particles and $N-1$ the sum of the pair forces between the particles.

Few years later, and thanks to the development of computational power and new numerical algorithms, the N-body simulations experience an exponential improvement. Nowadays the algorithms most frequently used are: Particle Mesh (PM), Particle-Particle/Particle-Mesh ($P^3M$) and TREE codes.

- **Grid based Particle-Mesh (PM) algorithm.** It utilizes a mesh to produce the density and potential and it was first applied to cosmology for the previously mentioned works of Doroshskevich [34] and Klypin and Shandaring [35]. The force computation scales as $N\log(N)$, with N the number of grid cells, allowing one more order of magnitude with respect to the PP method. However, the spatial resolution is limited to the cell size. Due to the limitation of the spatial resolution, the internal part of the clustered objects is not well described, and it is necessary to add short range forces connecting the cells in the mesh. This led to the next method.

- **Particle-Particle/Particle-Mesh ($\mathbf{P^3M}$).** This algorithm was developed in the 80's by Efstathiou and Eeastwood [36, 37] combining the mesh structure for the large-scale forces with small scale particle-particle forces contribution. The main disadvantage of this method comes when the clustering increases making the small scale force contribution dominates. For this reason, this method has been improved including subgrids in the strong clustering areas (high density), firstly proposed by Couchman in early 90's [38].

- **TREE algorithms.** This algorithm developed by Barnes and Hut [39] reduces the number of particles that interact by subdividing consecutively a cube (node) in smaller cubes where each node contains a single particle or subnodes. The center of each node is the center of mass and the force is computed by walking the tree using a multipole expansion related to the relative distance of the particles and the size of the node. The main disadvantage of this method is the number of operations, the need of storage the hierarchical tree and the difficulty to parallelize it, even though it has been used to simulate cluster formation in parallel implementation [40].

Forty years ago, N-body simulations started with only few thousands particle. Today, they can be done with more than $10^9$ particles, in volumes of hundreds of Mpc with a really high resolution, as is the case of the Millenium Simulation [23], used in the *Euclid* Collaboration [41] or the MultiDark simulation [42] for the Three Hundred Project [43]. Although N-body

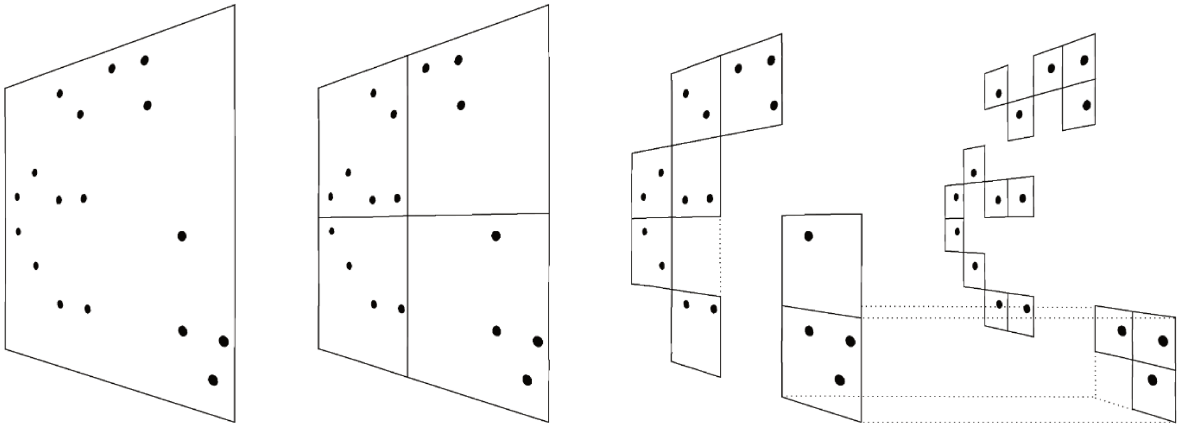

Table 1.4 – Barnes and Hut TREE representation in 2D. Credit: Springel et al. [44].

simulations has been fundamental for the understanding non-linear clustering and large scales, they are not a good approximation for studing the formation of galaxies. The information we get from the Universe comes from the baryonic matter, which makes difficult to compare observations with dark-matter only simulations. For this purpose a different approach was considered: the semi-analytical models (SAMs).

### 1.5.2 Semi-Analytical Simulations

Semi-analytical simulations appear in the 90's as an approximation to study galaxy formation. The gravitational part can be performed computing a probability function based on the Press-Schechter formalism and performing MonteCarlo realizations. For the baryonic physics, these simulations use physical models with free parameters that are compared with observations in order to calibrate them [45, 46, 47]. These models, though, are not enough to describe the morphology and structure of the galaxies, thus a different approach was proposed: combine N-body simulations, which describe large-scale structures and the cosmic web with high accuracy, with SAMs for the baryonic physics (gas dynamics and stars) [48]. With the increase of observations and the understanding of baryonic physics, star formation and feedback (effects of star formation on the interestellar medium) were included in these N-body/SAMs simulations [49, 50, 51]. They permit the construction of complete galaxy catalogues like the one used for the *Euclid* Collaboration [52]. This type of models are specially useful when a large sky area need to be simulated with high accuracy, as it is the case for the *Euclid* satellite [41], or ground based telescopes as The Rubin Observatory (LSST) [53].

### 1.5.3 Hydrodynamical Simulations

Computing realistic simulations is fundamental, to be able to make a direct comparison between simulations and observational data. Next step is to solve, both gravitational and gas dynamics equations at the same time because at small angular scales, baryons can modify the dynamics of the system since they are affected by other type of interactions like electromagnetic and strong interaction. This type of simulation is called Hydrodynamical simulations.

Hydrodynamical simulations appear in the late 80's and early 90's thanks to the improvement in the computing power [27]. They solve the gravitation and gas dynamics equations at the same time. There are two types of numerical methods used, depending on if the fluid elements are considered particles or meshes.

- **Smoothed Lagrangian Hydrodynamics (SPH).** This is a method where the fluid elements are represented by pseudo particles. It is the most popular one in astrophysics and cosmology and it was first proposed by Lucy [54] and Gingolg and Monnaghan [55] in 1977. The discrete fluid properties smoothed by a kernel function to reconstruct continous variables. This interpolation led to the same problem found in the TREE or $P^3M$ N-body methods: finding the nearest neighbors of the particle. Then, SPH is normally combined with TREE or $P^3M$ N-body simulations. The first cosmological code, written by Evrard in late 80's, combining $P^3M$ N-body with SPH [56]. Hernquist and Katz [57] a year after developed a combined TREE N-body with SPH and later more codes were delevoped for cosmology [58]. An example of this is the Gadget-X code developed by the Three Hundred Collaboration [43] that is based on the Gadget-2 TREE-SPH code [59]. More details about this can be found in Chapter 8.

- **Eulerian Methods.** This method treats the fluid elements as meshes. Thus the fluid equations are solved in a grid by finite-differences. The first Eulerian method used in cosmology was proposed by Cen et al in 1990 [60] combining their code with a PM/N-body code. They used the Godunov algorithm [61] that is a finite volume method. The spatial accuracy of this method can be improved when the gas dynamical quantities in a cell are not constant. When assuming a parabolic interpolation, the accuracy reaches third-order and the method is called Piecewise Parabolic Method (PPM) as proposed by Colella and Woodward [62] in the 80's and it is the standard method for cosmological eulerian models [63, 50, 64, 65]. Other Eulerian methods used in cosmology are: the Total Variation Diminishing (TVD) [66], Flux-Corrected-Transport (FCT) [67, 68], ZEUS-3D [69, 70] and Eulerian methods in non-uniform meshes [71].

Chapter 2

# Galaxy Clusters



After galaxies are formed, and due to the hierarchical nature of the growth of structures in the Universe, galaxy clusters are formed at a relatively low redshift, $z < 3$. They form in the intersection of filaments where there is a high concentration of dark matter. Clusters are the most massive gravitationally bound structures in the Universe. A typical galaxy cluster has a total mass between $10^{14} - 10^{15} M_{\odot}$ of which 85% corresponds to dark matter. A cluster is composed also of the intracluster medium (ICM) that is a hot ionised gas that comprises 12% of the cluster mass. Finally, the mass coming from the galaxy members is negligible, around 3%. The typical size of a cluster is the order of $R_{200} = 1 - 3$ Mpc, where $R_{200}$ is the radius from the center of the cluster at which the cluster mean density is 200 times the critical density of the Universe at the redshift of the cluster, i.e., $\Delta = \rho(r < R_{\Delta})/\rho_{crit}$ where $\Delta = 200$. Clusters are cosmological laboratories which can be studied at several wavelengths, and they are connected to the cosmological parameters.

In this chapter we describe cluster observables at different wavelengths. We also present some cluster observational properties in the optical/infrared that we will use during the thesis. These observational properties play a key role in cluster detection using the so called cluster finders. We introduce several cluster finders and how to compute the probability of finding a cluster at a certain mass and redshift. Finally, we describe the relationship between clusters and the cosmological parameters.

## 2.1 Cluster Observables

The direct information that can be obtained from a cluster is coming from its observables. The latter are useful to estimate the cluster number counts, the selection function, and the cluster mass and redshift, among others. Depending on the wavelength range in which the clusters are observed, we find different types of observables. Some of these observables, and surveys that inferred them, are described below.

### Milimetrical Observables

At milimetre wavelenghts we are sensitive to the intra-cluster medium (ICM) via the Sunyaev-Zel'dovich (SZ) effect. In this regime we find the NIKA2 camera [72], the *Planck* Satellite [73], the ACT (Atacama Cosmolohy Telescope, [74]) ot the SPT (South Pole Telescope, [75]), among others. The thermal Sunyaev-Zel'dovich (SZ) effect is a distortion of the CMB spectrum due to inverse Compton scattering the CMB photons with the hot electrons in the ionized ICM [76, 77]. For quantifying this effect we define the Compton parameter, $y$. It gives a measure of the electron pressure of the ICM along the line of sight. The Compton parameter integrated in an sphere of radius $R_{500}$ is defined as $Y_{500}$. The latter can be related to the cluster mass when considering hydrostatic equilibrium. The scaling relation between $M_{500}$ and $Y_{500}$ has been measured by the Planck Collaboration [73]:

$$Y_{500} = 10^{-4.19\pm0.02} E(z)^{2/3} \left[\frac{(1-b)M_{500}}{6x10^{14}M_{\odot}}\right]^{1.79\pm0.08} D_A^{-2} Mpc^{-2}, \tag{2.1}$$

with $E(z)$ and $D_A$ are defined by equation 1.12 and equation 1.15. Catalogues of clusters from SZ observations can be used to constraint cosmological parameters [78, 79, 80, 81].

### X-Ray Observables

Clusters are observed in X-Ray via bremsstrahlung emission of the ICM electrons. Main cluster observations in X-Rays come from XMM-Newton Survey [82], eROSITA [83] or Chandra [84]. The main observables at this wavelength range are the total X-Ray luminosity, $L_X$, the X-Ray temperature, $T_X$, the cluster gas mass, $M_{gas}$ or the thermal energy, $Y_X$ (equivalent to Y) that is the product of cluster temperature and the gas mass. Considering the self-similar scenario for clusters, all the previous observables can be related to the cluster mass through

scaling relations [85]:

$$\begin{aligned} L_X &= M^{4/3} E(z)^{7/3}, \\ T_X &= M^{2/3} E(z)^{2/3}, \\ M_{gas} &= M, \\ Y_X &= M^{5/3} E(z)^{2/3}. \end{aligned} \tag{2.2}$$

### Optical/Infrared Observables

The cluster information obtained at optical/infrared wavelengths comes from the light of stars and galaxies ([86, e.g.]). Furthermore, it is possible to detect and study clusters via gravitational strong and weak lensing on background galaxies (see Chapter 3). Current optical surveys as KiDS [87] and SDSS [88] will be followed by the *Euclid* Satellite [41], and the Vera Rubin Observatory (LSST) [53]. Optical/infrared surveys are used to measure the cluster redshift either spectroscopically (e.g., *Euclid*) or using various photometric bands (Rubin and *Euclid*). From these observables various estimates of the cluster mass can be found:

- **Cluster Richness:** It is an estimate of the number of galaxies within a cluster radius at a certain magnitude limit, $m_L$. It must be corrected by extracting the contamination of field galaxies. It serves as a mass proxy through scaling relations such as [89]:

$$\ln N_{200} = (0.47 \pm 0.12)(\log M_{200} - 14.5) + 1.58 \pm 0.04, \tag{2.3}$$

  where $N_{200}$ is the number of galaxies within $R_{200}$.
- **Velocity Dispersions:** The line of sight radial velocity dispersions, $\sigma_\nu$, of cluster's galaxies are a cluster mass proxy under the assumption of dynamical equilibrium through the Jeans equation [90, 91] or by scaling relations with other mass proxies like the richness [92]:

$$\log \sigma_\nu = (0.30 \pm 0.04)(\log N_{200} - 1.5) + 2.77 \pm 0.01. \tag{2.4}$$

- **Cluster Luminosity:** It measures the total luminosity of the cluster. As a mass proxy, just the sum of the brightest galaxies could suffice. However, to compute the total cluster luminosity, an extrapolation must be done for magnitudes above a certain limit, $m > m_L$. This extrapolation is usually done by fitting a function to the magnitude distribution and then integration from $m_L$. The most used function is the Schechter luminosity function [93] (see next section for more details). Generally the mass-lumiosity ratio, M/L, scales as a power-law with the cluster mass $M/L \propto M^\alpha$ with $\alpha \simeq 0.2 \pm 0.1$ [86].

## 2.2 Cluster Properties in the Optical/Infrared

### Luminosity Function

The Luminosity Function (LF) is defined as the number of galaxies in a volume, with a certain magnitude. Its shape depends on the galaxy types and the environment. Therefore, luminosity functions differ between background (or field) galaxies and cluster's galaxies. For this reason, luminosity functions are a key property for cluster's detection in the visible and IR surveys.

The apparent magnitude of a detected galaxy is measured through the flux at a given spectral band, Q:

$$m_Q \propto -2.5 \log(f_Q). \tag{2.5}$$

For the AB system, the apparent magnitude is $m_{AB} = -2.5 \log(f_\mu) + 8.9$ [94] where the flux spectral density, $f_\mu$ is measured in Jansky (Jy). The absolute magnitude for a spectral band, R, is defined as

$$M_R = m_Q - \mu - K_{QR}, \tag{2.6}$$

where, $m_Q$ is the apparent magnitude, $\mu$ is the distance modulus defined as $\mu = -5(\log(\frac{D_L}{10\,\mathrm{pc}}) - 1)$ with $D_L$ the luminosity distance, defined by equation 1.16. The K-correction, $K_{QR}$, accounts for the conversion between observed and rest frame in-band photometric measurements [95, 96]. The distance modulus defines the magnitude of an object as it would be seen at a distance of 10 pc.

Luminosity functions have been modelled analytically through time. The most used model is called the Schechter function [93, 97] given by

$$\Phi(m) = 0.4 \log(10) \phi^* 10^{0.4(m^* - m)(\alpha+1)} \exp\left(-10^{0.4(m^* - m)}\right), \tag{2.7}$$

with $\phi^*$, the normalization, $\alpha$, the faint-end slope, and $m_*$ the characteristic magnitude. An on-going problem in the computation of the luminosity function is that the Schechter model can not reproduce simultaneously both the bright and faint part of the luminosity profile.

### Galaxy Density Radial Profile

The inner structure of clusters is sensitive to the cosmological parameters [98] and it is a key property for the performance of cluster finders. It is important for a correct estimate of the cluster masses as well as for the understanding of the evolution and formation of the cluster

components [99, 100]. The most known galaxy number density profile model is the Navarro-Frenk-White profile [101], which describes clusters with a high central density. It is given by

$$n(r/R_{200}) = \frac{n_0}{(cr/R_{200})(cr/R_{200}+1)^2}, \tag{2.8}$$

with c the concentration, and $n_0$ the normalization. The relationship between the concentration and the cluster mass can be used as a cosmological probe [102]. The concentration is also a mass proxy. It has been shown that $c \propto M^{-\beta}$ with $\beta(z)$ [103].

Another profile model that has been proven to better fit CDM density profiles than the NFW profile is the Einasto profile [104, 105, 106]

$$\rho(r) = \rho_0 \exp\left(\frac{-2}{\alpha}\left[\left(\frac{r}{r_0}\right)^{\alpha} - 1\right]\right), \tag{2.9}$$

with $\rho_0$, $r_0$ and $\alpha$ are the free parameters of the model. The latter depends with the mass and redshift [103]. Notice that these profiles describes the mass density profiles, and during our thesis we will use them to describe the galaxy radial density distribution.

Other type of profiles have been proposed through time, depending on the cluster properties. For example, when studying the galaxy rotation curves of clusters, they show a central core [107]. A suggested profile for this type of cluster is the Burkert profile [108, 109]. Other typical profiles are the Hernquist [58] (similar to NFW in the inner part) and the isothermal sphere profile [110, 111].

The long discussion of how to model properly the galaxy distribution of clusters is still a fundamental topic, for the cluster mass determination, for galaxy formation and cluster formation and for cluster detection. Therefore, during this thesis we will use numerical simulations for modelling the galaxy density distribution within clusters.

## 2.3 Cluster Detection in the Optical/Infrared

To use galaxy clusters as cosmological probes we need a cluster sample, beside their mass estimation. Cluster catalogues are extracted from galaxy catalogues in which a cluster finder has been run. Therefore, the first thing is to construct a galaxy catalogue. Galaxy survey catalogues provide information about sky positions, redshift estimations via photometry or spectroscopy, luminosities, or richness estimations, among others. In the following we describe the main cluster properties that can be inferred from a galaxy catalogue that can impact the performance of a cluster finder.

Although clusters can be detected in several wavelengths as discussed above, historically the first cluster detections have been made via their galaxy members, in the optical and infrared domain [112]. There are several ways of detecting clusters depending on the information available. The first method for identifying and classifyng clusters was developed by Abell in the late 50's [113], characterizing them by their richness, i.e., the number of galaxies in a cluster in a magnitude range. He used apparent magnitudes to calculate distances and physical sizes for clusters. Abell created a cluster catalogue with approximately 4000 clusters. The main problem of Abell's cluster catalogue is that it is not complete and it is contaminated. Thus, a proper computation of the selection function is crucial for understanding the properties of the detected clusters. Nowadays clusters are searched for using cluster finders. Here we present a brief review of several cluster detection algorithms by the optical/infrared properties of clusters (for a more detailed review see [86]).

### Cluster Finders

- When the redshift is not available in the galaxy sample the most used algorithm is the Matched Filter (MF [114]). This method filters galaxies which do not belong to clusters. However, to construct the filter, one has to assume a form for the galaxy density radial profile and the luminosity function of cluster members. The idea is to perform a maximum likelihood estimator (see Chapter 4 for details). The theoretical model has two contributions, one from the background galaxies and the other from the cluster's galaxies

$$D(R, m) = b(m) + \lambda \Sigma(R) \phi(m), \tag{2.10}$$

  with $b(m)$ the background galaxy counts, $\phi(m)$ the luminosity function, the projected density radial profile, $\Sigma(R)$ and the richness $\lambda$. The likelihood will depend on the free parameters for the analytical models for $\Sigma(R)$ and $\phi(m)$. After filtering, clusters are detected looking for local maxima in a box of a given size centered in each galaxy. This method has been applied to several surveys like Sloan Digital Sky Survey (SDSS) [115] or others (e.g., [116, 117, 118]). In these catalogues the completeness reaches 100% for clusters with a mass $\sim 10^{14} M_{\odot}$ up to intermediate redshift, decreasing to 50% when increasing the mass and redshift.
- If the redshift information is available in the galaxy catalogue the most used identification algorithm is the Friends-of-Friends (FoF) percolation algorithm [119, 120]. In it, a volume is defined by two sizes, one along the sky plane, and another along the redshift (line of sight), centering the volume in a galaxy, and linking galaxies within this volume. It is

crucial for this method to choose a proper size and shape for the volume, having mainly two methods: from observational characteristics of the galaxies [121] or by using numerical simulations and observations [122]. This method has been applied to several surveys: SDSS [123], the Astrophysics Redshift Survey [124], among others (e.g., [121, 125]). In the output cluster catalogue, there are smaller clusters than the ones found with MF, of the order of $M \sim 10^{13} M_{\odot}$, meaning that projected spatial distributions does not allow to find massive cluster as does the 3D ones.

- The Cluster Red Sequence (CRS) [126], is a method based on another galaxy property: the color. Color is the difference between two photometric bands of the galaxy spectra, for example V-I. It has been found that cluster member galaxies are redder than the field ones, and also their morphology differs. The idea is to select a cut in the color-magnitude (V-I)-I diagram [126] to select galaxies whose redshift is close to the mean cluster redshift. In clusters, the oldest and brightest galaxies are placed in the center, which are also the more passive ones in terms of star formation. On the opposite hand, bluer galaxies are found in the outer part. At the same time, and paying attention to the morphology, elliptical galaxies are placed in the inner part on the clusters, while spirals in the outer part. The inner region of a cluster is denser, reason why it is possible to find more elliptical galaxies, that are formed from the accretion of the spirals ones. The CRS method looks for overdensities when computing the surface density of a galaxy color catalogue. This method is good to reduce galaxy field contamination. For redshifts $z > 1$, the red sequence become redder and deep IR measurements are necessary. For this reason, spaced-based telescopes play a key role as they can measure in the mid and far infrared (MIR, FIR), wavelengths which are absorbed by the Earth atmosphere. This method has found clusters up to $z \sim 2$ using the infrared Spitzer survey [127, 128], and other surveys [129, 130]. Another method using the red sequence for clusters is maxBCG [131], which uses the brightest cluster galaxy (BCG) for finding the cluster. It looks for a red bright galaxy which lives in an overdense region of galaxies whose color dispersion is small (red sequence). Applied to SDSS [132], the output was a catalogue with a purity of 90% and completeness of 85% for clusters with mass higher than $10^{14} M_{\odot}$.

- The Adaptative Matched Identifier of Clustered Objects (AMICO) algorithm [133], is a matched filter based finder. The filter is defined with a cluster model with a combination of a LF and a galaxy number density profile, and the noise with a spatially uniform LF. A 3D galaxy distribution is convolved with the filter and it generates a 3D amplitude map where each peak represents a detection. It has been applied to the synthetic galaxy catalogue used for the *Euclid* Collaboration [134] and they have found a completeness of

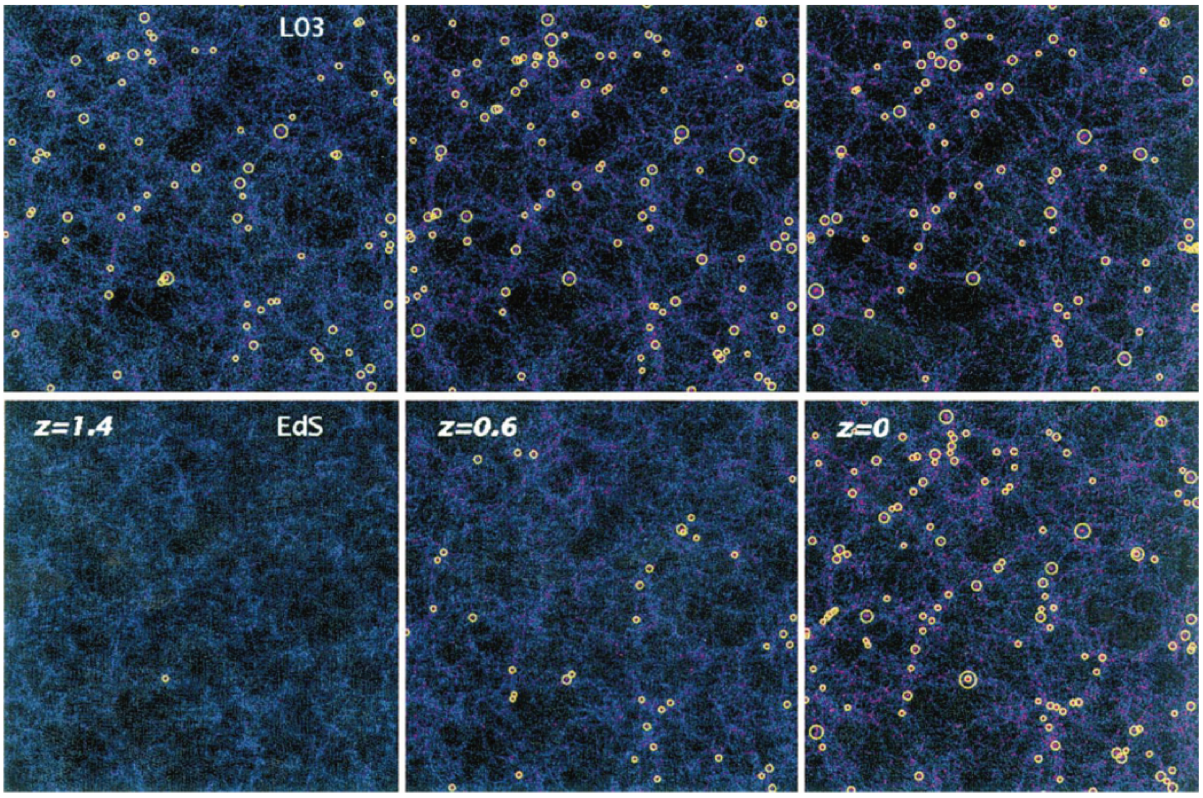


Figure 2.1 – N-body simulation for different cosmologies where the yellow circles represent clusters. The upper row represents a dark energy dominated Universe with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. The bottom row represents a matter dominated universe with $\Omega_m = 1$ and $\Omega_\Lambda = 0$. From left to right we show the redshift evolution for slices at $z = 1.4$, $z = 0.6$ and $z = 0$. Credit: Borgani and Guzzo [136].

85% and purity up to 95% for clusters of mass higher than $10^{14} M_\odot$ for a redshift $z < 2$. Furthermore, it has been recently applied to the KiDS survey data [135], obtaining a purity of 95% and a completeness up to 80% for clusters of mass higher than $10^{14} M_\odot$ for a redshift $z < 0.8$.

- The detection algorithm used in this thesis is PZWAV. It will be described in Section 7.2. It has been applied to the synthetic galaxy catalogue used for the *Euclid* Collaboration [134] and they have found a completeness of 80% and purity up to 90% for clusters of mass higher than $10^{14} M_\odot$ for a redshift $z < 2$.

## 2.4 Clusters as cosmological probes

### 2.4.1 Cluster Number Counts

The number of clusters as a function of mass and redshift is a major cosmological probe. As an example, Figure 2.1 shows a dark matter N-body simulation for two different types of cosmologies. As we can observe, the number of clusters, which are represented by yellow circles, as a function of redshift and mass, is different when considering a Universe dominated by dark energy with $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$ (upper row), or a matter dominated Universe with $\Omega_\Lambda = 0$ and $\Omega_m = 1$ (bottom row). The number of clusters as a function of the mass and redshift is given by:

$$\frac{dN}{dz} = \int d\Omega \int \chi(z, \hat{M}, a, b) \frac{dN}{dz dM d\Omega} dM, \tag{2.11}$$

where N is the number of clusters, $dN/dM$ is the Halo Mass Function (equation 1.20), $\Omega$ the solid angle, and $(z, M, a, b)$ are the redshift, cluster mass and sky position, respectively. The selection function, $\chi(z, \hat{M}, a, b)$, gives the probability of finding a cluster of redshift, $z$ and mass, $M$, at a position $(a, b)$. This function is an intrinsic characteristic of the survey and the detection algorithm that is used to detect clusters. The selection function is related to the completeness and purity of the detection algorithm and the data catalogue, i.e., the number of detected clusters (which we want to maximize) and the number of false positives in the detection (which we want to minimize). To estimate the selection function, in general, a threshold of detection is fixed from a criteria either in the purity or on the properties of the detected catalogue. For a given threshold and a given scaling relation between the mass proxy associated to the observable the probability of finding a cluster at a given mass and redshift is computed. In order to estimate the selection function, it is needed to use simulated data sets. It can be either a full mock galaxy catalogue including clusters and field galaxies, or individual simulated clusters which are injected in the observed galaxy catalogue. There has been proposed different ways of computing the selection function related to different levels of refinement and related to the cluster finder used [137, 135, 134, 138, e.g.].

The fact that clusters are formed from small early fluctuations, makes the number of clusters depending mainly on $\sigma_8$, and $\Omega_m$ [136]. However, clusters can also be used to constrain [139] the equation of state of dark energy, $w_{DE}$, and the growth factor, $D(z)$ [140]. The most challenging issues in cluster cosmology are the estimation of the mass, which is not a direct observable, and the estimation and improvement of the selection function.

The cluster mass can be inferred through several methods, for example: through the Jeans equation [90, 91] that uses the number of galaxies and velocity dispersions; through the consider-

ation of hydrostatic equilibrium with X-ray or SZ effect (explained below) measurements [141] ; through lensing effects [142] (gravitational lensing effect explained in Section 3.2); through scaling relations, where the mass is estimated by establishing a relation with a cluster observable such as, for example, the number of galaxies in a cluster (or richness) [143].

### 2.4.2 Other Cluster Comoslogical Probes

There are other ways of constraining cosmological parameters using cluster properties. In the following we review some of them.

- **Baryon Fraction:** The expected mass fraction of gas for a given cosmology of a halo at a redshift $z$ depends of cosmology through $D_A(z)^{3/2}$, and therefore, $\Omega_m$ and $\Omega_\Lambda$. By combining the observations with the theoretical prediction for a given cosmology it is possible to constrain $\Omega_m$ and $\Omega_\Lambda$ [144, 145].
- **Cluster Clustering:** The clustering of clusters can be used to measure baryonic acoustic oscillation via the correlation function or the power spectrum [146]. This probe by itself is not competitive with the others but combining with the cluster number counts leads to tighter constraints of the cosmological parameters [147]
- **Nature of Dark Matter:** Galaxy clusters have been one of the first cosmological objects from which it was inferred the existence of dark matter [18]. Numerical simulations predict that the dark matter distribution for relaxed clusters (spherical) can be expressed as a Navarro-Frenk-White (NFW) profile [101], whose inner density slope is $\rho_{DM} \propto r^{-1}$. For non-relaxed clusters this does not apply, neither when baryon physics is included, or for high resolution dark matter only simulations [106], finding that the inner cluster distribution follows a Einasto profile [104, 105]. Thus, dark matter has effects in galaxy cluster's internal structure, and they can constraint the nature of dark matter [148].

### 2.4.3 Current Cosmological Results

In current cosmological constraints the amplitude of the matter power spectrum, $\sigma_8$, and the matter density, $\Omega_m$, are degenerate. Thus, they are usually expressed as a combination defined as $S_8 \equiv \sigma_8\sqrt{\Omega_m/0.3}$. Figure 2.2 shows the actual constraints in $S_8$ obtained from different measurements: CMB (top part), Cluster counts (middle part), galaxy distribution (bottom part). The single vertical solid line represents the weighted mean over the low redshift cosmology, i.e., cluster counts and galaxy clustering. With respect to the CMB cosmology, there is a clear shift in the actual cosmological constraints. This can be due to new physics that changes the evolution of the Universe at high and low redshift or to observational and/or modelling systematics effects. For example in the case of clusters it could be related to uncertainties in the computation of the halo mass function, bias in the estimate of the cluster mass, misestimation of the selection

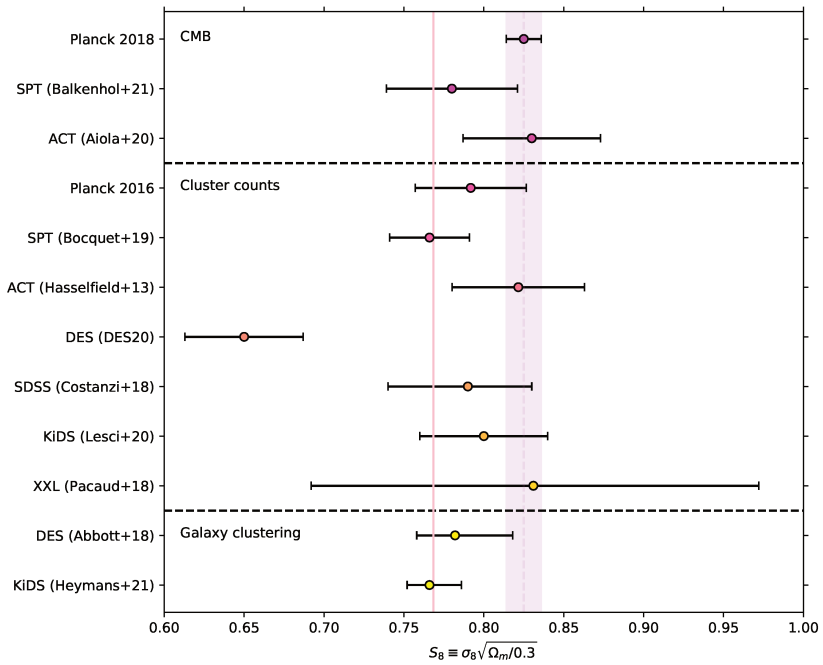


Figure 2.2 – Constraints in the cosmological parameter $S_8$ obtained from different measurements: CMB (top part), cluster counts (middle part), galaxy distribution (bottom part). From top to bottom the credits of the results are for: [4, 149, 150, 78, 79, 80, 81, 80, 151, 88, 87, 152, 81, 153]

function in catalogues, wrong modelling of the cluster physics. In this thesis we are going to focus in the cluster physics and in the estimation of the selection function, which can play an important role.

Part II

# The *Euclid* Mission

Chapter 3

# *Euclid*



In this chapter we present the *Euclid* mission. We start describing in general the satellite, then we focus on its instruments: the visible instrument, VIS, and the near infrared instrument, NISP. The latter is responsible for photometry and spectroscopy measurements that allow redshift estimations. These estimations will be done in a near infrared wavelength range that is difficult or impossible to access from ground based telescopes. Hence, *Euclid* is a key mission in the infrared domain. For this reason, we pay special attention to the NISP instrument, more precisely we describe its observing sequence and its infrared detectors. Next, we present the main scientific goals of the mission via the cosmological probes for which *Euclid* has been optimised. To conclude, we present the *Euclid* survey.

## 3.1 The Instruments

*Euclid* is a Medium Class mission of the European Space Agency's (ESA) Cosmic Vision 2015-2025 programme [41]. It is expected to be launched in 2023 on a Soyuz ST-2.1B rocket for its later insertion at the Second Sun-Earth Lagrangian Point, $L2$, 1.5 million kilometres away from the Earth. It will operate for 6 years. An artistic representation of the satellite is shown in Figure 3.1. The satellite is composed by a 1.2 meters Korsch telescope, with a field of view of 1.25x0.727 deg$^2$. The *Euclid* telescope is composed by three mirrors, the first one will collect the light to send it to the second mirror. The latter has a mechanism with three degrees of freedom that allows focus and tilt corrections. After the second mirror, the light will pass through several optical filters until the third and last mirror, that will direct the light flux to the

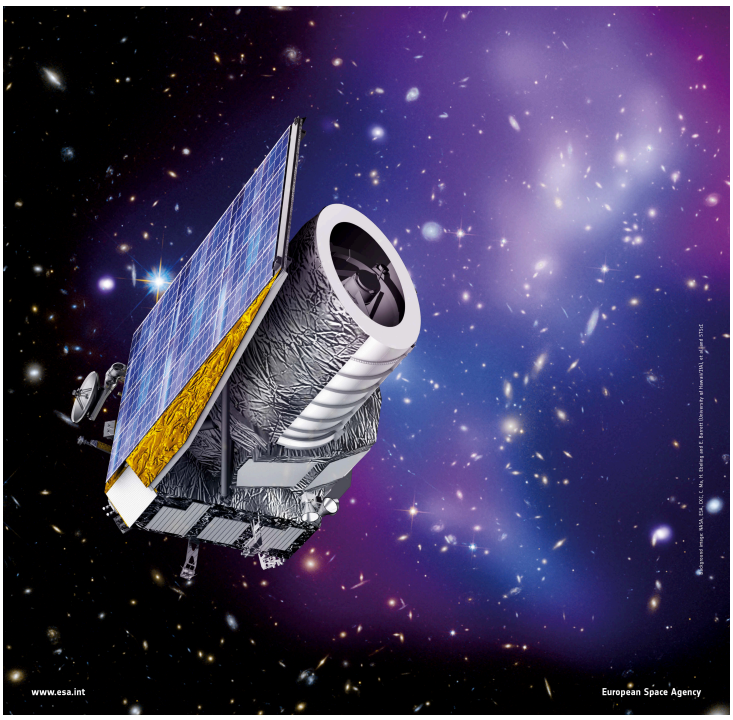


Figure 3.1 – Artistic representation of the *Euclid* satellite from ESA. Credit: https://www.euclid-ec.org/ .

satellite instruments: the visible instrument (VIS) for the reflected light, and the Near Infrared Spectrophotometer (NISP) for the transmitted light. In Figure 3.2 we show a general summary of the properties of the *Euclid* Satellite obtained from the *Euclid* definition study report [41]. In this chapter we will detail different aspects from this summary.

**Main Scientific Objectives**

***Understand the nature of Dark Energy and Dark Matter by:***

- Reach a dark energy *FoM* > 400 using only weak lensing and galaxy clustering; this roughly corresponds to 1 sigma errors on $w_p$ and $w_a$ of 0.02 and 0.1, respectively.
- Measure γ, the exponent of the growth factor, with a 1 sigma precision of < 0.02, sufficient to distinguish General Relativity and a wide range of modified-gravity theories
- Test the Cold Dark Matter paradigm for hierarchical structure formation, and measure the sum of the neutrino masses with a 1 sigma precision better than 0.03eV.
- Constrain $n_s$, the spectral index of primordial power spectrum, to percent accuracy when combined with Planck, and to probe inflation models by measuring the non-Gaussianity of initial conditions parameterised by $f_{NL}$ to a 1 sigma precision of ~2.

**SURVEYS**

| | Area (deg2) | Description |
|---|---|---|
| Wide Survey | 15,000 (required)<br>20,000 (goal) | Step and stare with 4 dither pointings per step. |
| Deep Survey | 40 | In at least 2 patches of > 10 deg$^2$<br>2 magnitudes deeper than wide survey |

**PAYLOAD**

| | | | | | |
|---|---|---|---|---|---|
| Telescope | 1.2 m Korsch, 3 mirror anastigmat, f=24.5 m | | | | |
| Instrument | VIS | NISP | | | |
| Field-of-View | 0.787×0.709 deg$^2$ | 0.763×0.722 deg$^2$ | | | |
| Capability | Visual Imaging | NIR Imaging Photometry | | | NIR Spectroscopy |
| Wavelength range | 550– 900 nm | Y (920-1146nm), | J (1146-1372 nm) | H (1372-2000nm) | 1100-2000 nm |
| Sensitivity | 24.5 mag<br>10σ extended source | 24 mag<br>5σ point source | 24 mag<br>5σ point source | 24 mag<br>5σ point source | 3 10$^{-16}$ erg cm-2 s-1<br>3.5σ unresolved line flux |
| Detector Technology | 36 arrays<br>4k×4k CCD | 16 arrays<br>2k×2k NIR sensitive HgCdTe detectors | | | |
| Pixel Size<br>Spectral resolution | 0.1 arcsec | 0.3 arcsec | | | 0.3 arcsec<br>R=250 |

**SPACECRAFT**

| | |
|---|---|
| Launcher | Soyuz ST-2.1 B from Kourou |
| Orbit | Large Sun-Earth Lagrange point 2 (SEL2), free insertion orbit |
| Pointing | 25 mas relative pointing error over one dither duration<br>30 arcsec absolute pointing error |
| Observation mode | Step and stare, 4 dither frames per field, VIS and NISP common FoV = 0.54 deg$^2$ |
| Lifetime | 7 years |
| Operations | 4 hours per day contact, more than one ground station to cope with seasonal visibility variations; |
| Communications | maximum science data rate of 850 Gbit/day downlink in K band (26GHz), steerable HGA |

**Budgets and Performance**

| | *Mass (kg)* | | *Nominal Power (W)* | |
|---|---|---|---|---|
| industry | TAS | Astrium | TAS | Astrium |
| Payload Module | 897 | 696 | 410 | 496 |
| Service Module | 786 | 835 | 647 | 692 |
| Propellant | 148 | 232 | | |
| Adapter mass/ Harness and PDCU losses power | 70 | 90 | 65 | 108 |
| **Total (including margin)** | **2160** | | **1368** | **1690** |

Figure 3.2 – Summary of the main properties of the *Euclid* satellite mission extracted from the *Euclid* Red Book [41].

### 3.1.1 The visible instrument - VIS

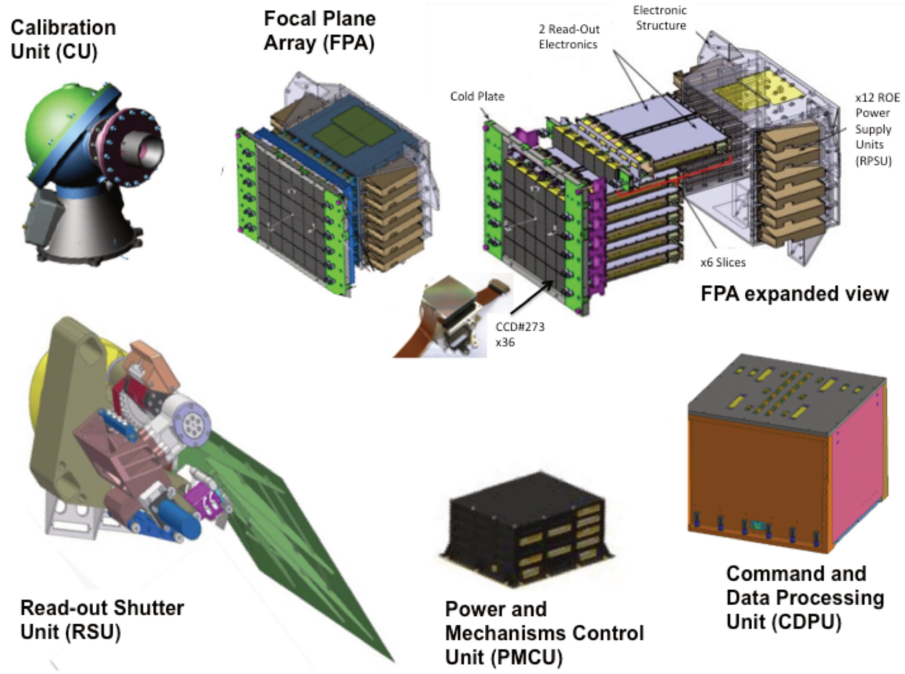


Figure 3.3 – Assembly of the VIS units. Credit: https://www.euclid-ec.org/.

The visible instrument (VIS) provides measurements of galaxy shapes via gravitational lensing with high quality images. It observes with a field of view (FoV) of 0.787x0.709 $deg^2$, 100 times bigger than the one of the Hubble telescope. The wavelength range covers a wide visible band from 550 to 900 nm. It is an assembly of several units, each one shown (excluding the VI-FH) in Figure 3.3:

- VIS Calibration Unit (VI-CU): Its purpose is to illuminate the focal plane with 12-LEDs panel for allowing flat fields for calibration.
- Read-out Shutter Unit (VI-RSU): Prevents the entry of light into the focal plane when performing readout operations or flat flied calibrations.
- Power and Mechanism Control Unit (VI-PMCU): It is dedicated to control all the VIS mechanisms as well as the calibration.
- Command and Data Processing Unit (VI-CDPU): It controls the data processing of the instrument.
- VIS flight harness (VI-FH): It connects the units.
- VIS Focal Plane (VI-FPA): It is composed of 36 arrays CCDs (charge-couple devices) of 4000x4000 pixels of 12$\mu m$ each one and a resolution of 0.1 arcsec. This unit is in charge of the detection of the visible light.

### 3.1.2 The near infrared instrument - NISP

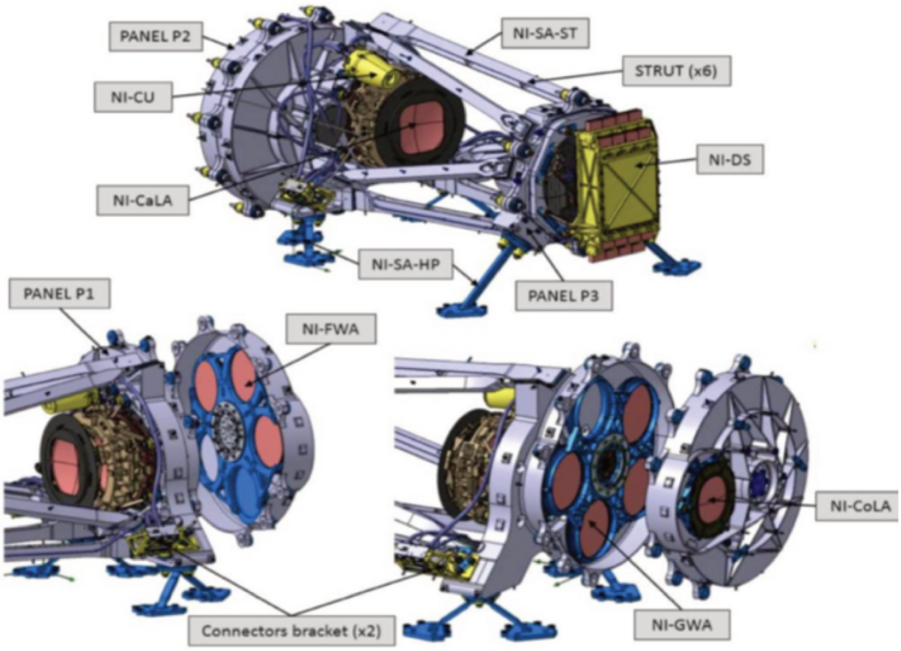


Figure 3.4 – Assembly of the NISP instrument. Credit: https://www.euclid-ec.org/.

The NISP instrument is a Near-Infrared Spectro-Photometer which allows to measure photometric and spectrometric redshifts of galaxies. The photometric mode has three filters $Y, J, H$ corresponding to wavelength ranging 920-1146 $nm$, 1146-1372 $nm$ and 1372-2000 $nm$, respectively. The spectroscopic mode performs in a wavelength range that goes from 1100 to 2000 $nm$ with a minimum flux sensitivity of 3 x $10^{-16}$ erg $cm^{-2}s^{-1}$. As shown in Figure 3.4, this instrument is composed of several subsystems of which the main ones are:

- NISP Calibration Unit (NI-CU) provides uniform illumination (flat-field) for calibration.
- NISP Grism Wheel Assembly (NI-GWA): A wheel that allows three different positions at 0°, 90° and 180°. This mechanism is fundamental for slitless spectroscopic measurements, where the spectra of each source (e.g., a galaxy) can overlap, hence, the angular discrimination is essential. This was the original design but has actually changed as there was a problem with one of the grisms. Now there are only two nominal red grisms with angles 4° and 184° plus a blue grism.
- NISP Filter Wheel Assembly (NI-FWA): A wheel containing five positions: the three photometric filters in the NIR, an opening and a close slot.
- NISP Warm Electronics (NI-WE): It is composed of the NISP Detector Control Unit (NI-DCU) and the NISP Data Processing Unit (NI-DPU) for the acquisition, processing and transfer of the data, and the NISP Instrument Control Unit (NI-ICU) which manage

the NI-GWA and NI-FWA mechanical system.

- NISP Detector system (NI-DS): It is the focal plane, with a FoV of 0.763x0.722 $deg^2$ and composed of 16 arrays of near-infrared H2RG detectors with 2kx2k pixels each, with a pixel resolution of 0.3 arcsec. More information about these detectors can be found in the following sections.

### H2RG Infrared detectors

The NISP detector system, NI-DS, contains the detector chain or Sensor Chip System (SCS). There the photons are captured and converted to electrons and to digital signal through the detectors in the Sensor Chip Array (SCA) and the electronics in the Sensor Chip Electronics (SCE). Both components are connected through the Cold Flex Circuit (CFC), as shown in Figure 3.5. The *Euclid* focal plane is composed of 16 H2RG detectors and their associated 16 SIDECARS ASIC.

In the SCA is located the chosen detector technology for *Euclid*, that are H2RG (Hawaii 2kx2k with Reference Pixels and Guide mode) detectors provided by Teledyne [154]. These detectors consist of HgCdTe pixel arrays with 2048x2048 pixels (see Figure 3.6), that collects the photons. Each array is composed of 32 sub-arrays of 64x2048 pixels with 32 parallel lecture channels. The $Hg_{1-x}Cd_xTe$ material has a tunable bandgap [155] that allows measurements in the NIR range for the right element proportion, $x$. The HgCdTe array is connected to a multiplexer that allows one to choose which pixels we want to read, or which ones we want to reset (set their value to zero). Therefore, the main advantages of the H2RG detectors for *Euclid* are: 1) high sensitivity in the NIR band ($0.9\mu m < \lambda < 2\mu m$), and, 2) being able to choose the pixels, and therefore the data, we want to process.

To control the readout electronics (multiplexer) of the SCA, Teledyne has developed an application specific integrated circuit (ASIC), the so-called SIDECAR (System for Image Digitization, Enhancement, Control and Retrieval) ASIC. One of the advantages of these circuits is that their noise is negligible when comparing with the H2RG readout noise, which dominates the system H2RG-SIDECAR. In the NISP context, this system is configured to have 32 channels that allow us to read in parallel the 32 arrays of 64x2048 pixels each one with a final output of an image of all the pixels. The total exposition time per frame is 1.445s where each pixel is measured in parallel. The SCE then is the responsible of the image readout and transfer (or not) to the Data Processing Unit of the NISP (NI-DPU). For each pixel the measured signal is obtained in a non-destructive way, that means we obtain the integrated signal per pixel.

In the next chapter, we are going to study the H2RG readout noise implication in the photon

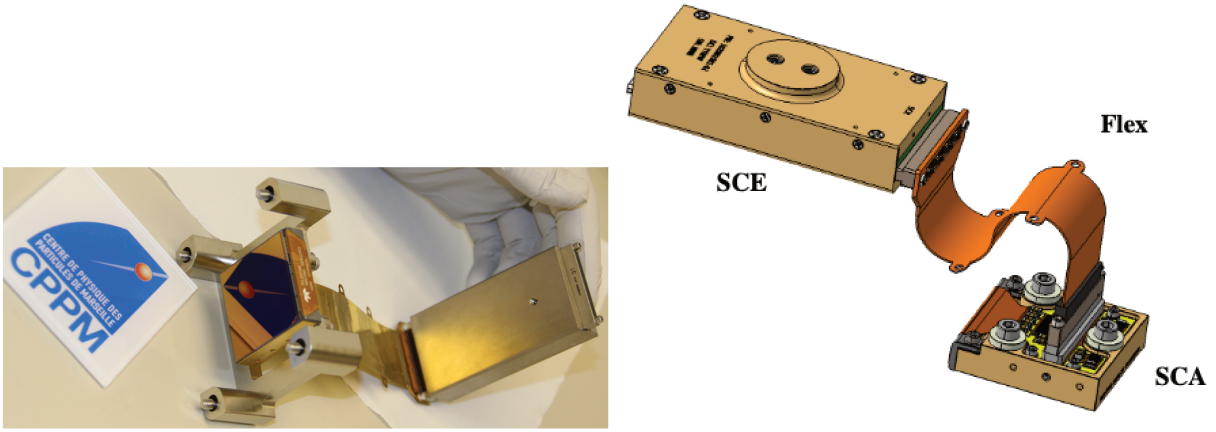


Figure 3.5 – NISP instrument Sensor Chip System (SCS). Credit: Aurelia Secroun at CPPM (left), Clémens et al. [156] (right).

flux measurement, presenting different readout modes for these detectors.

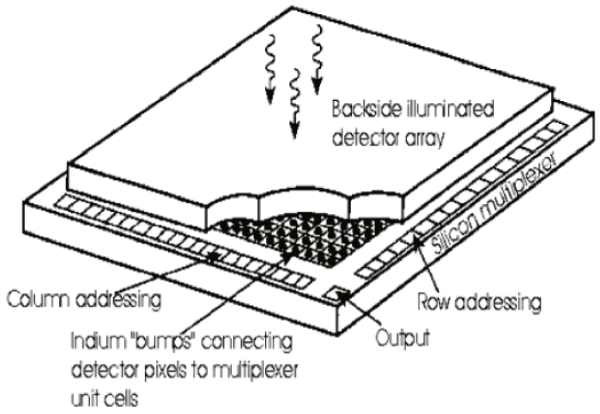


Figure 3.6 – Hybrid infrared detector. Credit: [157].

### NISP Observed Sequence

The NISP Grism Wheel Assembly (NI-GWA) allows spectroscopy measurements at the different angles with the red grism and with the blue grism. The NISP Filter Wheel Assembly (NI-FWA), includes the three photometric filters, as shown in the bottom left and bottom right part of Figure 3.7, respectively. The GWA had originally five positions: one open window, one blue filter, and three red filters that allow the rotation for avoiding the overlapping at GW1 = 0°, GW2 = 90° and GW3 = 180°. The FWA has also five positions, three of them corresponds to the three photometric filters Y,J,H, and the other two are the close and open positions. With these two wheels the NISP instrument completes a whole cycle of spectro-photometry observa-

tions as shown in Figure 3.7, before the problems with one of the red grisms. Currently only two red grisms of the four grisms are working in the wide field observations. To our understanding the current whole spectro-photometric cycle is done as follows:

- First three spectrometric modes are used:
  1. GWA1 with position at 0° and FWA open, observing for the duration of 574s.
  2. GWA2 with position at 184° and FWA open, observing for the duration of 574s.
  3. GWA3 with position at 4° and FWA open, observing for the duration of 574s.
  3. GWA4 with position at 180° and FWA open, observing for the duration of 574s.
- Then the three photometric modes:
  4. GWA with open position and FWA with filter $Y$, observing for the duration of 121s.
  5. GWA with open position and FWA with filter $J$, observing for the duration of 116s.
  6. GWA with open position and FWA with filter $H$, observing for the duration of 81s.
- For each spectro-photometric cycle, a region of 0.54 deg$^2$ is covered. Then the satellite is pointing to another sky region and a new cycle is started. During this time the last measure is done:
  7. FWA in close position. This is done for account on the sky background noise.

## 3.2 Main Scientific Goals

The *Euclid* satellite is mainly devoted to cosmology and intends to unveil the nature of Dark Energy, and Dark Matter. The goal is to measure the 3-dimensional distribution of matter of the Universe, for tracking the formation of structures, and having a better understanding of the expansion of the Universe and its acceleration. *Euclid* will focus mainly in the redshift range $0 < z < 2$ where the Large Scale Structures are formed. For this purpose, there are two primary cosmological probes that *Euclid* has been optimised for, the Weak Gravitational Lensing (WL) and Baryonic Acoustic Oscillations (BAO), and other secondary ones, such as Galaxy Clusters.

### Weak Gravitational Lensing

The path of the light that travels through the Universe can deviate due to the gravitation effects of massive objects, a phenomenon known as Gravitational Lensing. This phenomenon induces an effect of distortion in the shapes and size of the galaxies, known as cosmic shear [158], which has been first proved in 2000 by different groups simultaneously [159, 160, 161, 162]. The cosmic shear is related to the gravitational field of large-scale structures, thus to their formation and evolution through time. Therefore, also to the cosmological $\Lambda CDM$ model, and in particular, to the matter density and the amplitude of the matter fluctuations, $\Omega_M$ and $\sigma_8$, respectively. In general this effect is small, of the order of 1%, thus the importance of acquiring high quality images for this small distortion is essential. For this *Euclid* will detect billions of images of

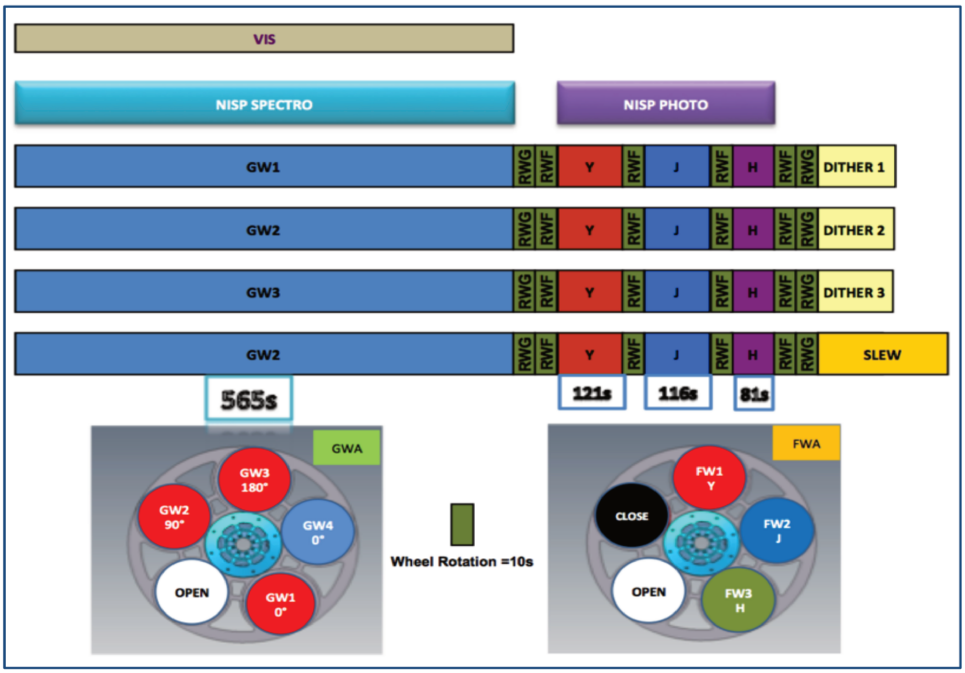


Figure 3.7 – NISP observing sequence. Credit: [163]

galaxies, observing in an unprecedented high quality. As an example, in Figure 3.8, we can see how the galaxy cluster Abell 2218, distorts the shape of the background galaxies. A key aspect in the determination of the cosmic shear, in addition to a high-quality estimate of the shape and size of the galaxy, is an accurate computation of the redshift of the galaxies. For this purpose, other surveys will be complementary to *Euclid*, like The Rubin Observatory [53]. *Euclid* will perform photometric redshift estimations in three photometric bands, while LSST in another different six photometric bands. However, the image quality of a ground based telescope is influenced by the atmosphere, for that both surveys are complementary to each other.

### Baryonic Acoustic Oscillations

The Universe, at the early stage, was a hot plasma and photons and baryons were coupled, the radiation pressure produces small oscillations in the baryons, compressing and expanding the plasma. The struggle between the radiation and gravity produces acoustic oscillations, which frozen after the decoupling of light and baryons, and are imprinted in the CMB (see Chapter 1 for a CMB description). These small perturbations, described by oscillatory solutions of a sound wave propagation, are called Baryonic Acoustic Oscillations (BAO). Due to these small perturbations, galaxies evolution suffer from what is knows as galaxy clustering, that means that galaxies are not randomly distributed in the Universe. The current distribution of galaxies is given by the BAO through the measurement of the two point correlation function (2PCF)

Figure 3.8 – Distortions of background galaxies due to the massive Abell 2218 galaxy cluster. Image by Hubble Space Telescope [Image credit: NASA, ESA, Richard Ellis (Caltech) and Jean-Paul Kneib (Observatoire Midi-Pyrenees, France)]

expressed as follows:

$$\psi(r) =< \delta(x)\delta(x+r) >, \tag{3.1}$$

where $\delta$ is the density contrast.

The 2PCF gives us the space correlation between two galaxies one at position, x, and the other displaced a distance, r. As an example, in Figure 3.9 from [164], we can see the two point correlation function (2PCF) over the comoving distance [1]. The magenta line shows a pure CDM model without BAO, showing a significant bias. This Figure tell us that the distribution of galaxies is correlated, then BAO are related to the space distribution of galaxies through time, and their position give us valuable information. These acoustic perturbations are small, affecting mainly large-scale physics. From the 2PCF we can see that the computation of the distance between objects is crucial, and therefore, the estimation of the position of each one. Since the signal is small, high accuracy in the determination of the distance is fundamental, for that *Euclid* will estimate galaxy redshifts between $0.7 < z < 2.1$ perfomig spectroscopy. The BAO as cosmological probe is then mainly related to the expansion of the Universe (Hubble constance $h$), the matter density $\Omega_m h^2$ and the baryon density $\Omega_b h^2$ [165].

### Galaxy Clusters

*Euclid* will be particularly well-adapted to detect cluster of galaxies. The number of clusters as a function of mass and redshift as their spatial distribution constitute major cosmological

1. Comoving distances are defined as the physical parameter divided by the scale factor $a(t)$, hence comoving distances are constant in time in an expanding Universe.

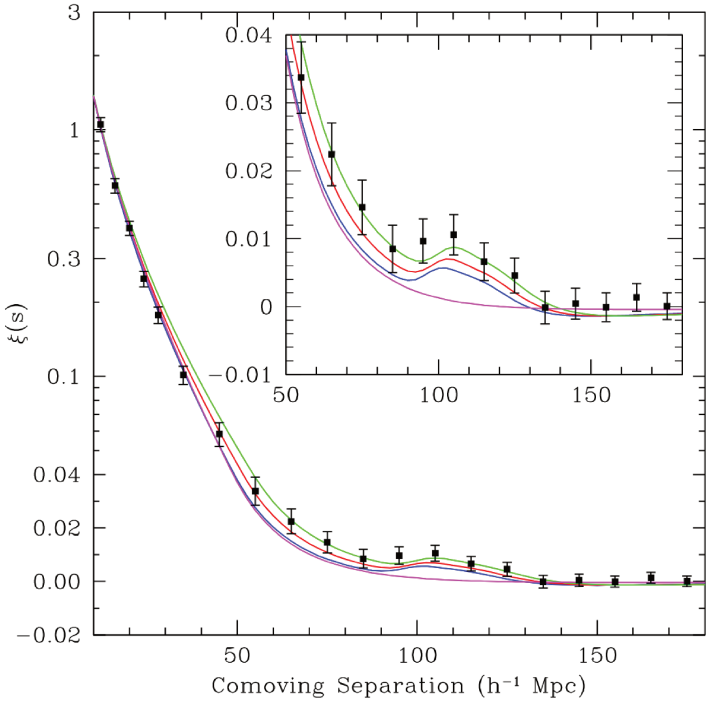


Figure 3.9 – Correlation function of galaxies from the SDSS survey by [164]. The bump at $100h^{-1}$Mpc corresponds to a peak of the barionic acoustic oscillations (BAO).

probes. *Euclid* will detect around $10^5$ galaxy clusters and will be a key instrument for cluster cosmology [137]. See Chapter 1 for more details about galaxy clusters as cosmological probes.

## 3.3 Expected Cosmological Constraints from *Euclid* Galaxy Clusters.

The forecast constraints on cosmological parameters from the expected galaxy cluster samples of *Euclid* are presented in Sartoris et al. [137], in great detail. Here we just describe the most interesting results.

Figure 3.10 shows the number of clusters above a given redshift to be detected with overdensities $N_{500,c}/\sigma_{field}$ in the *Euclid* photometric survey (dotted blue and solid red lines, respectively). Where $N_{500,c}$ and $\sigma_{field}$ are the cluster member galaxies within a radius of $R_{500}$, and the field galaxies rms. The histograms show the number density of clusters expected to be detected within redshift bins of width $\Delta z = 0.1$ for the same detection thresholds (dotted cyan and solid magenta histograms, respectively). *Euclid* will detect a total number of clusters of about 2 x $10^5$ at a threshold of $N_{500,c}/\sigma_{field} > 5$, and about 4 x $10^5$ at a treshold of $N_{500,c}/\sigma_{field} > 3$.

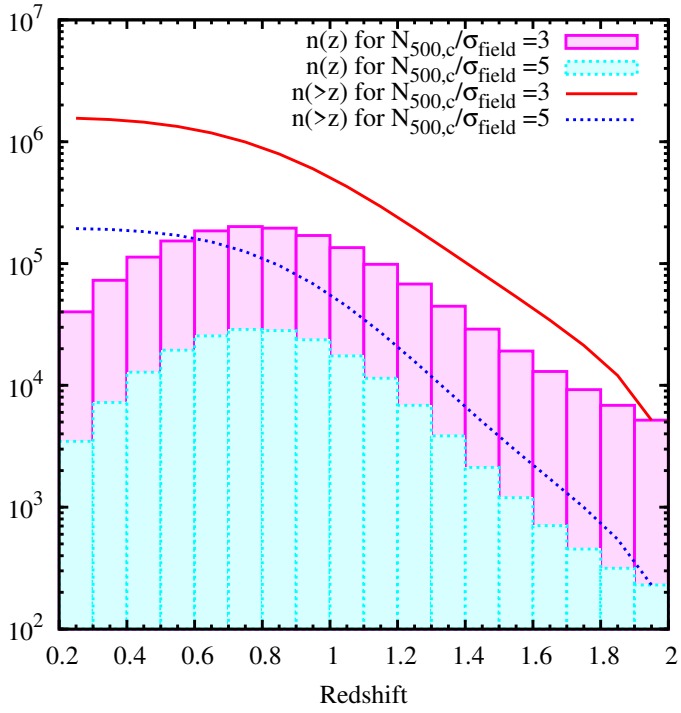


Figure 3.10 – Number of clusters above a given redshift to be detected with overdensities $N_{500,c}/\sigma_{field}$ in the *Euclid* photometric survey (dotted blue and solid red lines, respectively). The histograms show the number density of clusters expected to be detected within redshift bins of width $\Delta z = 0.1$ for the same detection thresholds (dotted cyan and solid magenta histograms, respectively). Figure adapted from Sartoris et al. [137].

Figure 3.11 shows the forecasted constraints on cosmological parameters from *Euclid* photometric clusters on the most interesting pairs of cosmological parameters. The ellipses correspond to 68 C.L. In each of the figures, the blue dotted contours are obtained by the number counts (NC) fisher matrix (FM) and the cluster power spectrum (PS) FM (Sartoris et al. [137, see]) assuming no prior information on any cosmological or nuisance parameters, for a cluster sample with a selection of $N_{500,c}/\sigma_{field} \geq 3$. The green dashed-dotted contours are obtained in the same way but adding strong priors on the scaling relation between the true and observed clusters mass, labelled as '+known SR'. The magenta solid contours have been obtained accounting for

*Planck* CMB constraints, labelled as '+*Planck* prior'. The cyan solid contours represent the same information of the magenta solid lines but for a cluster sample with a selection of $N_{500,c}/\sigma_{field} \geq 5$, labelled as $5\sigma$. The yellow solid curve shows the constraints combining NC and PS FM with *Planck* prior but without information on the nuisance parameters. The expected sensitivity of *Euclid* to the rms fluctuations at 8 $h^{-1}$ Mpc scales is $\Delta\sigma_8 = 0.0014$, to the mass density, $\Delta\Omega_m = 0.0011$ and with respect to dark energy, $\Delta\omega_0 = 0.03$ and $\Delta\omega_a = 0.2$. For $\Omega_m$ and $\sigma_8$ we expect an improvement in a factor of five in the uncertainties with respect to the *Planck* 2018 combined CMB results [4]. We expect an improvement of a factor of nine and three for $\omega_0$ and $\omega_a$, respectively. The optical/infrared survey KiDS, has detected 7899 clusters up to redshift 0.8 for a threshold SNR $> 3.5$ [135]. In compare with future surveys in the optical/infrared, LSST will detect about $10^5$ clusters up to redshift 1.4, and it will help to constraint the cosmological parameters with a similar precision to *Euclid* [166]. Thus, *Euclid* will represent a key experiment for cluster cosmology and cluster physics in the next decades.

## 3.4 Survey

After its launch, *Euclid* will orbit the $L2$ Sun-Earth lagrangian point during 6 years while performing its survey. It will observe billions of galaxies mainly in the range $0 < z < 2$. For these redshifts, 30% of the light is invisible from ground, and for the remaining light, bright night sky lines dominate the background. Therefore, *Euclid* provides a pioneering contribution thanks to its photometry and spectroscopy in the NIR wavelength for this redshift range.

The coverage by the *Euclid* survey is illustrated in galactic coordinates map in the upper part of Figure 3.12. The center of the map corresponds to the Milky Way, this galactic plane region is avoided as galactic emission will contaminate distant galaxies. Moreover, the blue regions are the wide survey that covers 15000 $deg^2$ up to a magnitude of 24.5 for VIS and 24 for NISP. In addition, the yellow areas are three Deep Fields where *Euclid* will perform a deep survey that covers 40 $deg^2$ up to about two magnitudes deeper than the wide survey. These areas consist of: 1) The *Euclid* Deep Field North (EDF North), covers 10 $deg^2$, found in the top left part of the Figure, 2) the *Euclid* Deep Field Fornax (EDF Fornax) located in the bottom right part, cover a 10 $deg^2$ sky region and 3) the *Euclid* Deep Field South (EDF South) that covers 20 $deg^2$. The latter is particularly interesting as it will be the first time this region is covered by a deep survey. Both the deep field found in the north part of the map, and the other south deep fields are the closest as possible to the Ecliptic Poles. This will allow a maximum coverage throughout the year.

The bottom part of Figure 3.12 shows the individual observed fields where each patch represents the shared field of view by VIS and NISP of 0.54 $deg^2$. Each color corresponds to a full

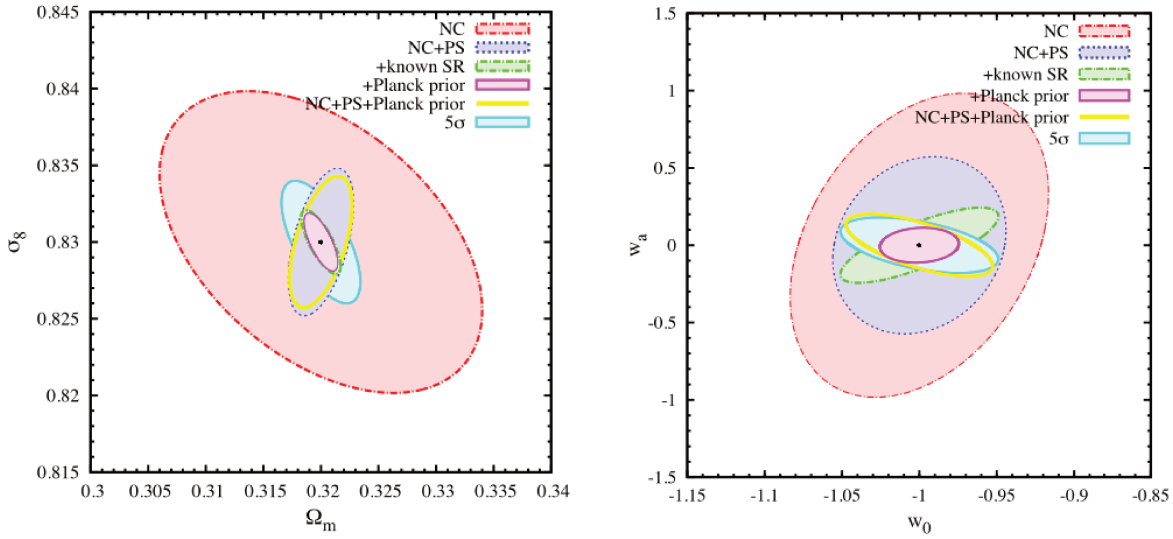


Figure 3.11 – Constraints at the 68 percent C.L. on the parameters $\Omega_m$ and $\sigma_8$ (left-hand panel) and on the parameters $\omega_0$ and $\omega_a$ for the dark energy equation of state evolution (right-hand panel). In each panel shows the forecasts for the $N_{500,c}/\sigma_{field} \geq 3$ *Euclid* photometric cluster selection obtained by (I) NC, the FM NC (red dash-dotted contours), (II) NC+PS, the combination of FM NC and PS information (blue dotted contours), (III) NC+PS+known SR, i.e., by additionally assuming a perfect knowledge of the nuisance parameters (green dash-dotted contours), and (IV) NC+PS+known SR + *Planck* prior, i.e., by also adding information from *Planck* CMB data (magenta solid contour). The yellow solid curve shows results from $N_{500,c}/\sigma_{field} \geq 3$ sample in the case NC+PS+*Planck* prior, i.e., with no assumption of the nuisance parameters. The cyan solid lines show forecasts for $N_{500,c}/\sigma_{field} \geq 5$ *Euclid* photometric cluster selection in the case NC+PS+known SR + *Planck* prior (labelled $5\sigma$). *Planck* information includes prior on $\Lambda CDM$ parameters and the dark energy equation of state parameters. Figure adapted from Sartoris et al. [137].

year of observation. The isolated fields over the sky are calibration fields. Observations will be performed in a step-and-stare mode: both VIS and NISP complete the observations in a field before moving to another one. As this, *Euclid* measures an area between 10 to 20 deg$^2$ per day and, therefore, patches of about 400 deg$^2$ per month. Finally, each 6 months, the satellite will point in the opposite direction to be able to measure the other hemisphere. After 6 years, the *Euclid* survey will provide the shape and photometric redshift of about 1.5 x 10$^9$ galaxies and the spectroscopic redshift of about 5 x 10$^7$ galaxies.

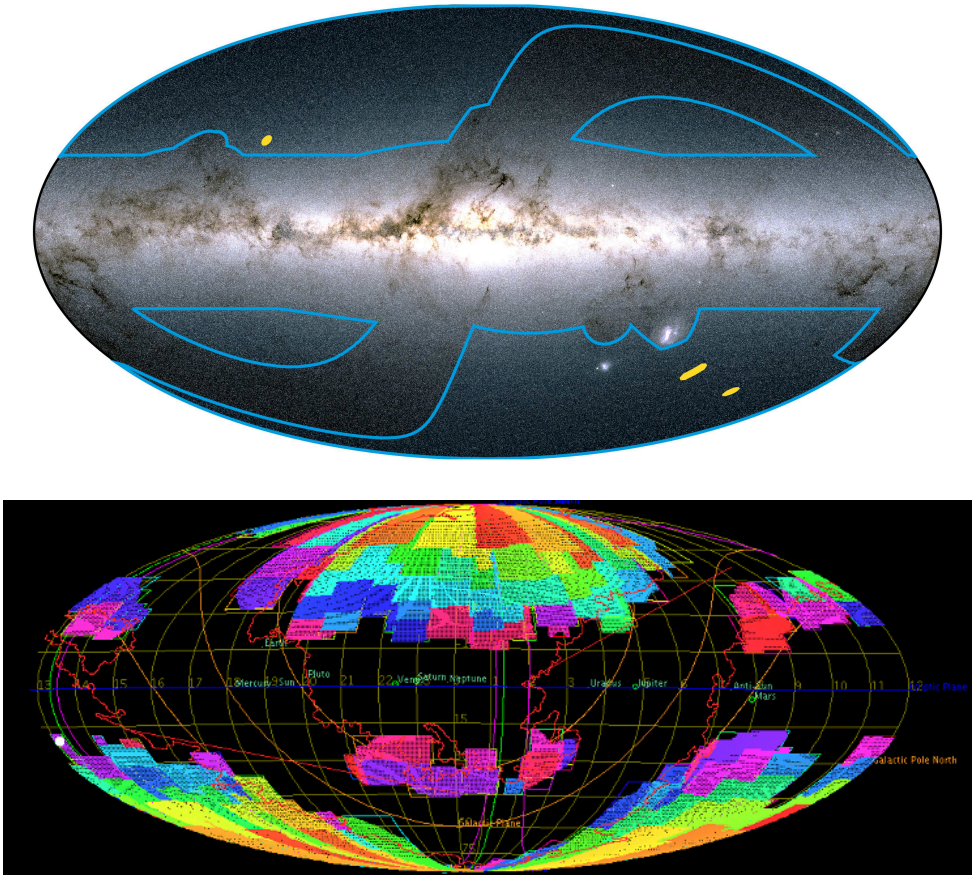


Figure 3.12 – *Euclid* Coverage Map. Credit: https://www.cosmos.esa.int/web/euclid/euclid-survey (upper part). https://www.euclid-ec.org/?page_id=2581 (bottom part).

Chapter 4

# Implication of correlated readout noise for flux measurement with the *Euclid* NISP instrument



In this chapter we are going to introduce the implication of correlations in the readout noise of the *H2RG* detectors, presented in Section 3.1.2, when the NISP instrument performs photon flux measurements in flight.

## 4.1 Readout Modes

As explained in Section 3.1.2 the infrared detectors used for *Euclid* perform non-destructive measurements of the photon flux in a pixel, for every pixel. We are going to describe briefly the different methods for the readout mode of the detector.

### 4.1.1 Up the Ramp (UTR)

In this mode the data is acquired in a non-destructive way, that means that the signal is accumulated through the different measurements. The acquisition starts by a reset that removes any previously accumulated signal. Then, the image from the detectors is read and transferred to acquisition system in a regular time steps. In Figure 4.1, the red lines are the reset values, where the signal is set to zero, and the blue ones are where the signal is acquired. Every single measure, represented in the figure by a vertical coloured line, is called frame. We define $t_{frame}$ as the acquisition time for a frame. The total exposure time is expressed as $t_{expo} = M \cdot t_{frame}$

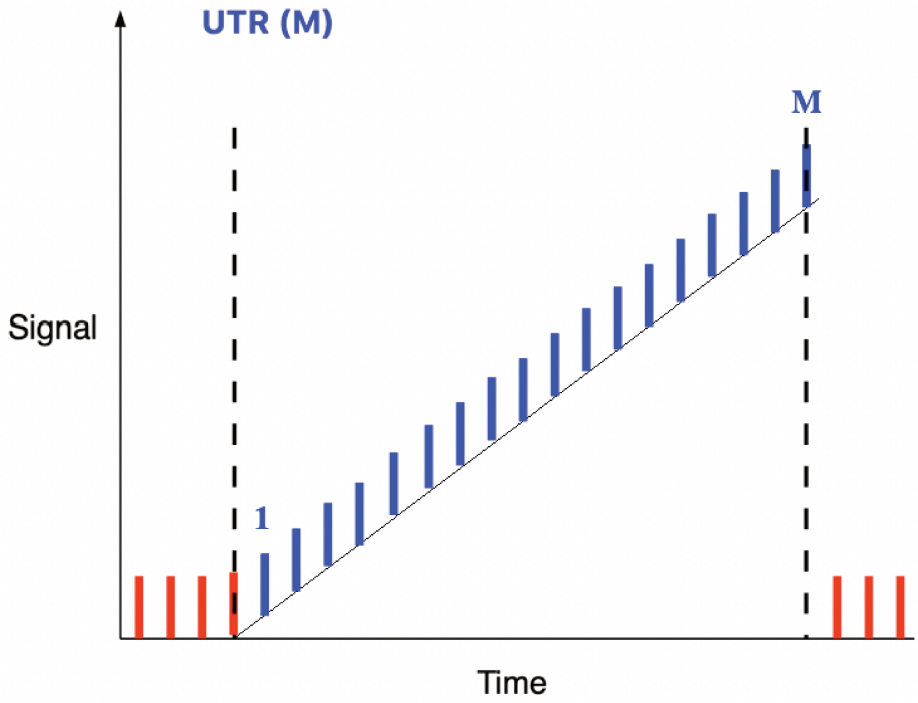


Figure 4.1 – Sketch of the Up-The-Ramp (UTR) Lecture mode.

where $M$ is the total number of frames.

In terms of noise, there is a contribution from the photon flux and another one from the readout process, that is intrinsic to the detector. The readout noise scales as $1/\sqrt{M}$ where $M$ is the number of acquired measurements. Hence, more acquired data means an important reduction of the readout noise.

The final flux in this method can be obtained by a linear fit of the ramp, since the cumulative behaviour comes from non-destructive exposures of the incident photons. Even though this method will give us an accurate estimation of the flux, the high amount of data transferred is not compatible with the in-flight processing limitations.

### 4.1.2 Correlated Double Sample (CDS)

The Correlated Double Sample (CDS) readout mode measures the photon signal two times: at the beginning of the exposure, $S_i$ and at the end, $S_f$, after a exposure time $t_{expo}$. As an example, Figure 4.2 shows a representation of the CDS readout mode. The red lines are the reset values, where the signal is set to zero, the blue ones are where the signal is measured and transferred, and for the green ones, the signal is not acquired. The final value is given by the difference of the signal $\Delta S = S_f - S_i$, in both measurement, where $S(t)$ is the measured signal at a time t, in analog-to-digital (ADU) units. Its uncertainty, $\sigma_{CDS}$, is related to the noise coming from the photon flux, and the readout noise intrinsic to the detector.

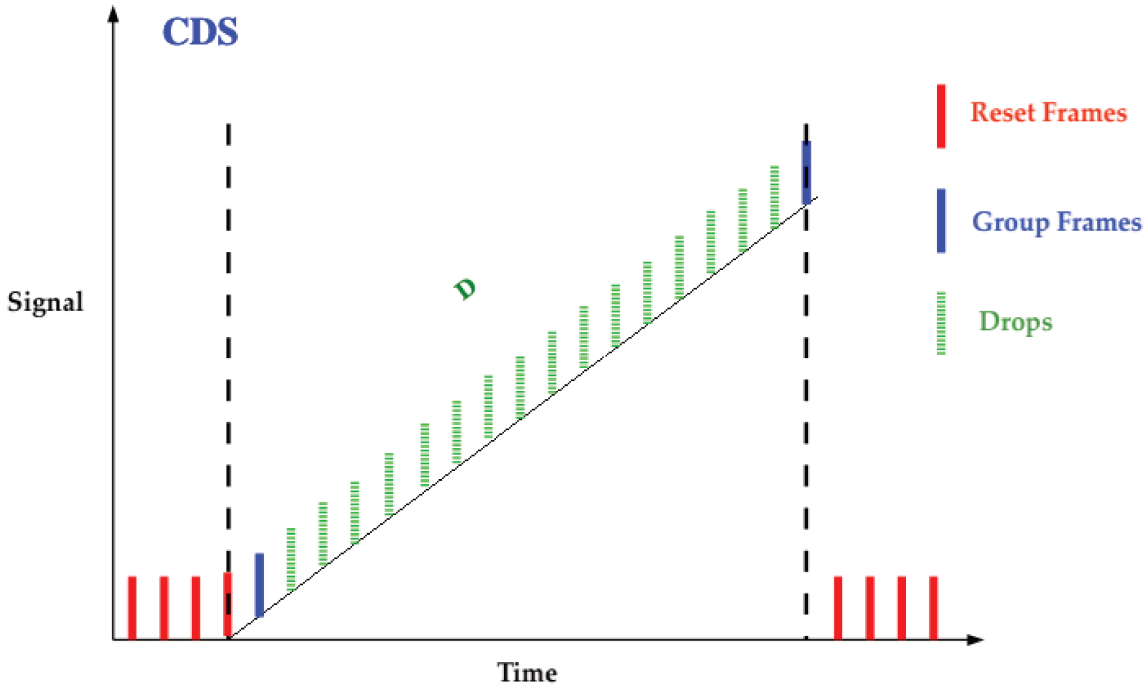


Figure 4.2 – Sketch of the Correlated Double Sample (CDS) readout mode.

### 4.1.3 Fowler-M

As explained in Section 4.1.1, a simple way of reducing the readout noise is to combine the data from various frames. The Fowler-M [167] readout mode read and transfer consecutive measures at the beginning and at the end of the exposure. Each set of consecutive acquired measurements is called a group. Then for the Fowler-M readout mode we have two groups, $G_1$ and $G_2$, as shown in Figure 4.3. In this case the readout noise is scaled by $1/\sqrt{M}$, where $M$ represents the number of frames in each group. The flux can be estimated by the difference of the total accumulated signal in each group, divided by the exposition time between the groups. Ideally, for a good estimation of the flux, we would like to have more measurements in the middle of the ramp, what leads us to the next readout mode.

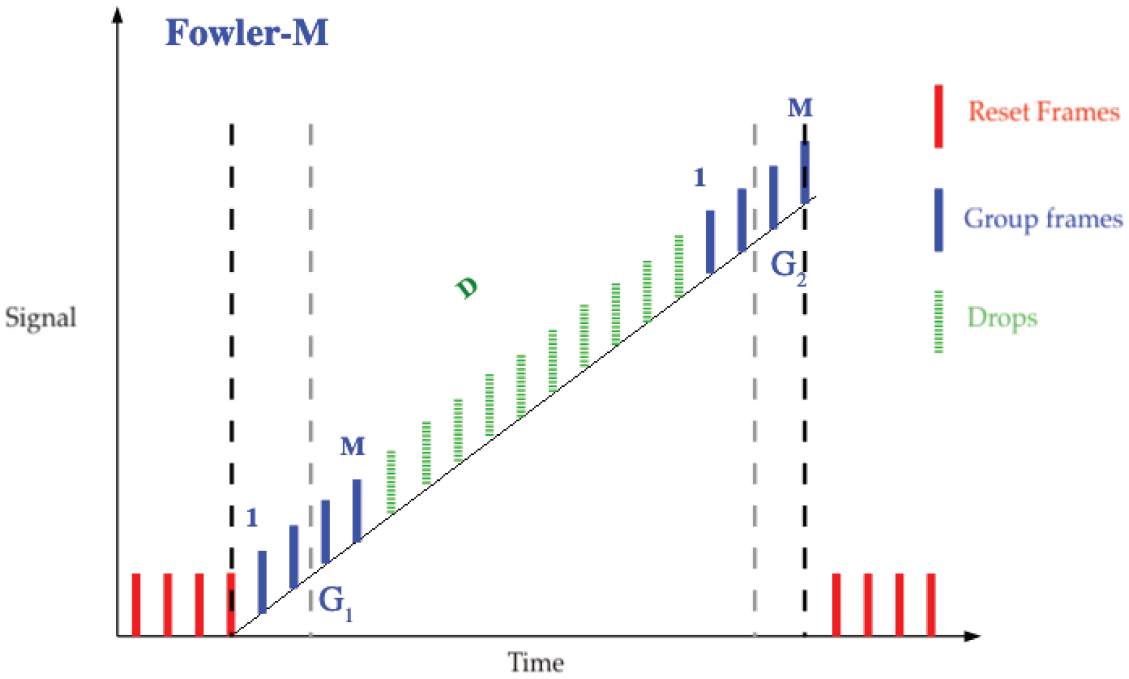


Figure 4.3 – Sketch of the Fowler-M readout mode.

### 4.1.4 Multiple Accumulated (MACC)

This mode is a compromise with respect to the previous ones. In order to reduce the readout noise, several frames will be acquired, but not all of them, for the in-flight CPU being able to manage the data. As shown in Figure 4.4, we read and transfer consecutive measurements in groups, as in the Fowler-M method, separated by some frames that are not transferred, called drops. We define $n$ as the number of groups, with $m$ frames each, and $d$ is the number of dropped frames in the $n-1$ drops. The notation for this readout mode is usually MACC[n,m,d]. This is the method that will be used in the in-flight configuration of the NISP instrument.

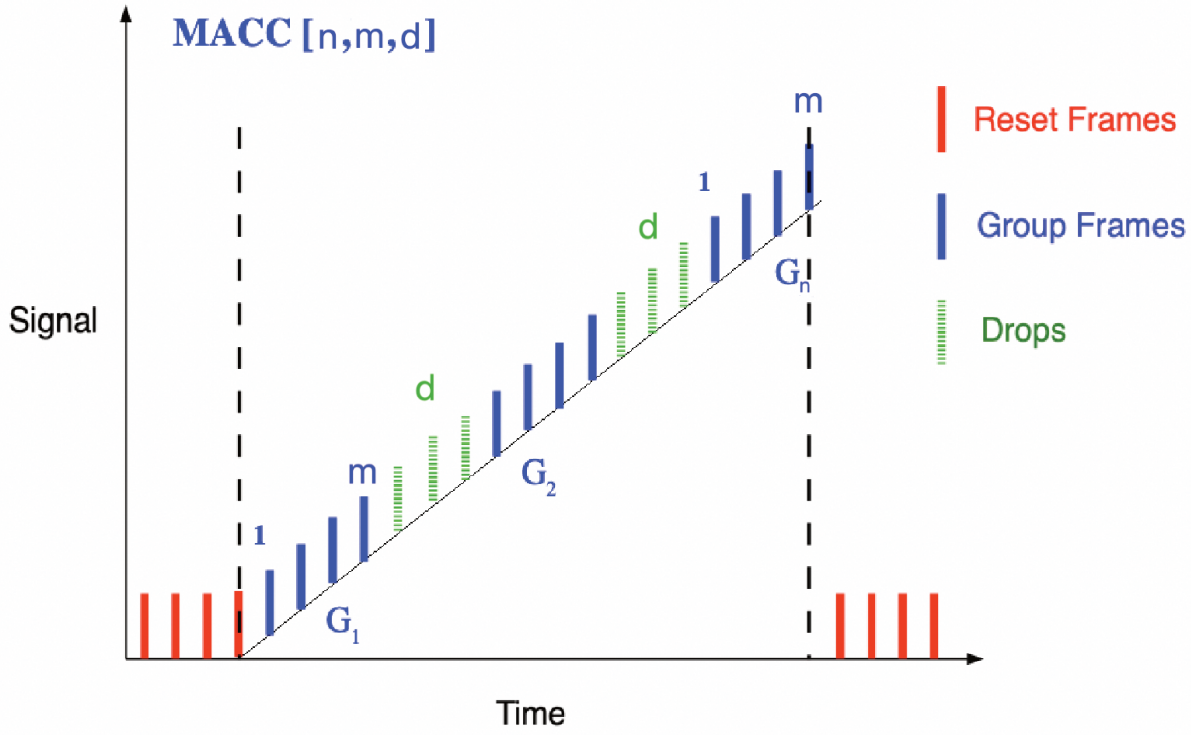


Figure 4.4 – Sketch of the Multiple Accumulated, MACC, readout mode.

## 4.2 Flux estimation in MACC readout mode

Due to on-board CPU limitations, the flux is determine from the slope from a linear fit to the ramps of the MACC readout mode. As a consequence to obtain an accurate flux estimate it is necessary to have an accurate description of the photon and readout noise. In the current *Euclid* baseline both the readout noise is described by a white noise approximation [167, 168]. The main properties of the readout noise assuming white noise are characterized during ground calibration and used in-flight. However, it has been found that individual H2RGs may present some level of correlated noise in the form of $(1/f)^{\alpha}$-like noise [169, 170]. Such noise correlation might bias *Euclid* in-flight flux estimates.

Here we concentrate on the MACC readout mode, used in-flight. Following [168] we estimate the total flux from the group differences, $\Delta G_k = G_{k+1} - G_k$ since the uncertainties in the process are lower than the estimation from the single groups. For one group $G_1$, the signal of consecutive frames is given by:

$$\begin{aligned} S_1^{(1)} &= \rho_1^{(1)} + f_0 \\ S_2^{(1)} &= \rho_2^{(1)} + f_0 + f_1^{(1)} \\ S_3^{(1)} &= \rho_3^{(1)} + f_0 + f_1^{(1)} + f_2^{(1)} \\ &\dots \\ S_m^{(1)} &= \rho_m^{(1)} + f_0 + \dots + f_{m-1}^{(1)} \end{aligned} \tag{4.1}$$

where $S_m^{(1)}$ represents the total signal at the last frame of the first group, $f_{m-1}$ is the photon flux signal that is accumulating over all the group frames[1], and $\rho_m^{(1)}$ is the readout noise of the last frame. In the second group we have:

$$\begin{aligned} S_1^{(2)} &= \rho_1^{(2)} + f_0 + \dots + f_{m-1}^{(1)} + \mathcal{D}^{(1)} \\ S_2^{(2)} &= \rho_2^{(2)} + f_0 + \dots + f_{m-1}^{(1)} + \mathcal{D}^{(1)} + f_1^{(2)} \\ S_3^{(2)} &= \rho_3^{(2)} + f_0 + \dots + f_{m-1}^{(1)} + \mathcal{D}^{(1)} + f_1^{(2)} + f_2^{(2)} \\ &\dots \\ S_m^{(2)} &= \rho_m^{(2)} + f_0 + \dots + f_{m-1}^{(1)} + \mathcal{D}^{(1)} + f_1^{(2)} + \dots + f_{m-1}^{(2)} \end{aligned} \tag{4.2}$$

where a new term is added, $\mathcal{D}^{(1)}$, representing the non-acquired accumulated signal of the drops between the groups $G_1$ and $G_2$. Accounting for signal and readout noise contributions, $G_k$, the

1. The notation following [168] defines $f_0$ as the first photon flux signal between the last reset and the first frame. Thus, the last frame photon flux notation is defined by $m-1$

averaged measured signal for group $k$, is given by:

$$G_k = \frac{1}{m}\sum_{i=1}^{m} S_i^{(k)} = \frac{1}{m}\sum_{i=1}^{m} \rho_i^{(k)} + \frac{1}{m}\sum_{i=1}^{m-1}(m-i) f_i^{(k)} + f_0 + \sum_{j=1}^{k-1}\left[\sum_{i=1}^{m-1} f_i^{(j)} + \mathcal{D}^{(j)}\right], \tag{4.3}$$

and then the signal for the group differences is:

$$G_{k+1} - G_k = \mathcal{D}^{(k)} + \frac{1}{m}\sum_{i=1}^{m-1}\left[i f_i^{(k)} + (m-i) f_i^{(k+1)}\right] + \frac{1}{m}\sum_{i=1}^{m}\left(\rho_i^{(k+1)} - \rho_i^{(k)}\right). \tag{4.4}$$

### 4.2.1 Maximum Likelihood Estimator

Using the group differences, we can then derive the total flux, $g$, by a linear fit. In [171] was noticed that the slope estimation by the typical Least Squares Fitting (LSF) fit method with equally weighted errors can be improved when considering correlated and uncorrelated errors. For that reason the method we will use from now on for the estimation of the accumulated flux is the Maximum Likelihood Estimator.

Let's consider a random variable $x$, that it is associated to a measurement as a way of describe it, that can take different possible numerical values $x_1, x_2, x_3...$, corresponding to different possible outcomes. The corresponding probabilities $P(x_1), P(x_2), P(x_3)...$ form a probability distribution, and $P(x)$ is called the Probability Density Function (PDF). Now, we consider a probability function that depends on a parameter $a$ (a particular realization) and the set of $N$ independent random variables $x$, $f_a(x_i) = f(x_i; a)$. The full PDF is given by:

$$f_a(x) = \prod_{i=i}^{N} f_a(x_i), \tag{4.5}$$

and the function that actually depends on the parameter $a$ is called the Likelihood Function [172], $L(a)$. As an example, equation 4.6, defines the probability of the outcome value $x_1$ to be observed when the true value of the parameter is $a$. This parameter $a$, can be multidimensional.

$$L(a|x_1) = f_a(x = x_1). \tag{4.6}$$

Under the hypothesis of the parameter $a$ being close to the true value, we expect a maximum

probability of finding it, then the likelihood function should be maximal for this parameter:

$$\frac{\partial}{\partial a}L(a) = 0, \tag{4.7}$$

and this condition is used to find the parameter, $a$. Let's consider now the case where the probability function is a Gaussian distribution, and we want to estimate its mean value, $\mu$. For practical reasons we consider the logarithmic of the likelihood. Then, the Gaussian PDF for a variable $x_i$ (supposing that the PDF is the same for every $x_i$, like the uncertainties, $\sigma$), is given by:

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right), \tag{4.8}$$

and the logarithmical likelihood of the whole set of random variables is expressed as:

$$L(\mu) = -\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^2} + N\ln\frac{1}{\sqrt{2\pi}\sigma}. \tag{4.9}$$

From this expression we can define the so known chi-square, $\chi^2$, function:

$$\chi^2 = \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^2}. \tag{4.10}$$

It gives information about how far from the true value our data are, or what is the same, how good our fit is.

As explained before, for estimating the mean value, $\mu$, the derivative of the likelihood should be $\frac{\partial}{\partial\mu}L(\mu) = 0$. Now, let's consider the 2D Gaussian PDF for two set of random variables, $x, y$, each one with its uncertainties, $\sigma_x, \sigma_y$, and mean values, $\mu_x$,$\mu_y$, given by:

$$f(x,y) = f_x{\cdot}f_y = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\exp\left(-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} + 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right). \tag{4.11}$$

Another way of writing this expression is using the error matrix, M, that contains the information about the uncertainties, $\sigma_x$ and $\sigma_y$ and the discrepancy vector, $X$. This matrices can be written as follows:

$$M = \begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_y\sigma_x & \sigma_y^2\end{pmatrix}. \tag{4.12}$$

$$X = \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}. \tag{4.13}$$

The final 2D Gaussian PDF can be written as:

$$f(x,y) = \frac{1}{2\pi|M|^{1/2}} \exp\left(-\frac{1}{2}X^T M^{-1} X\right) = \frac{1}{2\pi|M|^{1/2}} \exp\left(-\frac{1}{2}\chi^2\right), \tag{4.14}$$

where $\chi^2 = X^T M^{-1} X$ and |M| is the determinant of the error matrix. This matrix is also called the covariance matrix and it can be written in a general way as:

$$M = \begin{pmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{pmatrix}, \tag{4.15}$$

defining the covariance $cov(x,y)$ as the expectation value between the two random variables $(x,y)$, and given by:

$$cov(x,y) = E[(x-\mu_x)(y-\mu_y)] =< x,y >, \tag{4.16}$$

$\mu_x$ being the expectation value of the variable $x$, also written as $E[x]$. This matrix give us information about the correlation between these two random variables.

Now we can generalize to multi-dimensional Gaussian distribution in matrix notation as follows:

$$f(x_1, x_2, ..., x_N) = \frac{1}{2\pi|M|^{1/2}} \exp\left(-\frac{1}{2}X^T M^{-1} X\right), \tag{4.17}$$

being the covariance matrix:

$$M = \begin{pmatrix} cov(x_1,x_1) & cov(x_1,x_2) & ... & cov(x_1,x_N) \\ cov(x_2,x_1) & cov(x_2,x_2) & ... & cov(x_2,x_N) \\ ... & ... & ... & ... \\ cov(x_N,x_1) & cov(x_N,x_2) & ... & cov(x_N,x_N) \end{pmatrix}, \tag{4.18}$$

or

$$M = \begin{pmatrix} < x_1,x_1 > & < x_1,x_2 > & ... & < x_1,x_N > \\ < x_2,x_1 > & < x_2,x_2 > & ... & < x_2,x_N > \\ ... & ... & ... & ... \\ < x_N,x_1 > & < x_N,x_2 > & ... & < x_N,x_N > \end{pmatrix}. \tag{4.19}$$

Coming back to the flux estimation for group differences explained in section 4.2, we want to

fit the slope of the MACC readout mode signal, given by $\Delta G$, to estimate the flux $g$, using the previous maximum likelihood estimator. We use the following Gaussian approximation for the likelihood function:

$$L = \frac{1}{\sqrt{2\pi|M|}} exp\left[-\frac{1}{2}(\Delta G - g)M^{-1}(\Delta G - g)^T\right], \tag{4.20}$$

where $\Delta G = \{\Delta G_k,\ k = 1, n-1\}$ is a vector gathering all group differences. $M$ is the covariance matrix for the group differences, and $|M|$ is its determinant. Next step is to determine the covariance matrix for the group differences in order to be able to use the Maximum Likelihood Estimator.

## 4.3 Analytical Covariance Matrix

As explained in the sections 4.1 and 4.2, the estimation of the flux is coming from the fit of the slope of the MACC readout mode. This fit will be done by using the so called Maximum Likelihood Estimator, but previously the covariance matrix should be computed. We discuss here the computation of the group noise covariance matrix, $C$, and the group difference noise covariance matrix, $D$, in the case of white readout noise (see [168]) and in the case of correlated readout noise. With respect to the photon noise, the flux integrated over a frame is Poisson distributed and stochastically independent between frames. This applies both to fluxes of frames within a group and within a drop. We can then write the covariance as:

$$\begin{aligned} < \delta f_i^k \delta f_j^l > &= f\delta_{ij}\delta_{kl}, \\ < \delta \mathcal{D}^k \delta \mathcal{D}^l > &= \mathcal{D}\,\delta_{kl}. \end{aligned} \tag{4.21}$$

### 4.3.1 White Noise Case

The white readout noise is assumed to be Gaussian distributed with a constant width $\sigma_R$ and zero mean as explained in [168]. Thus, we can write:

$$< \delta\rho_i^k \delta\rho_j^l >= \sigma_R \delta_{ij}\delta_{kl}. \tag{4.22}$$

Using the definition of a group in 4.3, the stochastic fluctuation that is associated are:

$$\delta G_k = \frac{1}{m}\sum_{i=1}^{m}\delta\rho_i^{(k)} + \frac{1}{m}\sum_{i=1}^{m-1}(m-i)\delta f_i^{(k)} + \delta f_0 + \sum_{j=1}^{k-1}\sum_{i=1}^{m-1}\delta f_i^{(j)} + \sum_{j=1}^{k-1}\delta\mathcal{D}^{(j)}. \tag{4.23}$$

Then, the group noise covariance matrix using the definition of covariance matrix 4.19 is given

by:

$$\begin{aligned} C_{kk}^{white} &=< \delta G_k \delta G_k > \\ C_{kl}^{white} &=< \delta G_k \delta G_l >, \end{aligned} \tag{4.24}$$

where $C_{kk}^{white}$ and $C_{kl}^{white}$ are the white noise approximation group noise covariance matrices for the diagonal and off-diagonal terms, respectively, and $\delta G_k$ the stochastic fluctuations of the $k$ group. From the group fluctuations we obtain:

$$\begin{aligned} C_{kk} =< \delta G_k \delta G_k >= \Bigg\langle \Bigg[ \sum_{p=1}^{k-1} \sum_{i=1}^{m-1} \delta f_i^{(p)} + \sum_{p=1}^{k-1} \delta \mathcal{D}^{(p)} + \frac{1}{m} \sum_{i=1}^{m} \delta \rho_i^{(k)} + \frac{1}{m} \sum_{i=1}^{m-1} (m-i) \delta f_i^{(k)} + \delta f_0 \Bigg] \\ \times \Bigg[ \sum_{q=1}^{k-1} \sum_{j=1}^{m-1} \delta f_j^{(q)} + \sum_{q=1}^{k-1} \delta \mathcal{D}^{(q)} + \frac{1}{m} \sum_{j=1}^{m} \delta \rho_j^{(k)} + \frac{1}{m} \sum_{j=1}^{m-1} (m-j) \delta f_j^{(k)} + \delta f_0 \Bigg] \Bigg\rangle \end{aligned} \tag{4.25}$$

$$\begin{aligned} C_{kl} =< \delta G_k \delta G_l >= \Bigg\langle \Bigg[ \sum_{p=1}^{k-1} \sum_{i=1}^{m-1} \delta f_i^{(p)} + \sum_{p=1}^{k-1} \delta \mathcal{D}^{(p)} + \frac{1}{m} \sum_{i=1}^{m} \delta \rho_i^{(k)} + \frac{1}{m} \sum_{i=1}^{m-1} (m-i) \delta f_i^{(k)} + \delta f_0 \Bigg] \\ \times \Bigg[ \sum_{q=1}^{l-1} \sum_{j=1}^{m-1} \delta f_j^{(q)} + \sum_{q=1}^{l-1} \delta \mathcal{D}^{(q)} + \frac{1}{m} \sum_{j=1}^{m} \delta \rho_j^{(l)} + \frac{1}{m} \sum_{j=1}^{m-1} (m-j) \delta f_j^{(l)} + \delta f_0 \Bigg] \Bigg\rangle \end{aligned} \tag{4.26}$$

That after some algebra gives [see 168, for details]:

$$\begin{aligned} C_{kk}^{white} &= (k-1)\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)(2m+1)}{6m} + \frac{\sigma_R^2}{m}, \\ C_{kl}^{white} &= (k-1)\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)}{2}. \end{aligned} \tag{4.27}$$

We can repeat this process to compute the group differences covariance matrices:

$$\delta G_{k+1} - \delta G_k = \delta \mathcal{D}^{(k)} + \frac{1}{m} \sum_{i=1}^{m-1} \left[ i \delta f_i^{(k)} + (m-i) \delta f_i^{(k+1)} \right] + \frac{1}{m} \sum_{i=1}^{m} \left( \delta \rho_i^{(k+1)} - \delta \rho_i^{(k)} \right). \tag{4.28}$$

and so the group difference noise covariance matrix, $D$, is given by:

$$\begin{aligned} D_{kk}^{white} &=< \delta(G_{k+1} - G_k) \delta(G_{k+1} - G_k) > \\ D_{kl}^{white} &=< \delta(G_{k+1} - G_k) \delta(G_{l+1} - G_l) >, \end{aligned} \tag{4.29}$$

being $D_{kk}^{white}$ and $D_{kl}^{white}$ are the white noise approximation group noise covariance matrices for

the diagonal and off-diagonal terms, respectively. We repeat the same process as for the group noise covariance matrix (further details in Appendix A of [168]), and we obtain:

$$\begin{aligned} D_{kk}^{white} &= \mathcal{D} + f\frac{(m-1)(2m-1)}{3m} + \frac{2\sigma_R}{m}, \\ D_{kl}^{white} &= \frac{f}{6m}(m^2-1)\delta_{(k+1)l} - \frac{\sigma_R^2}{m}. \end{aligned} \tag{4.30}$$

### 4.3.2 Correlated Readout Noise Case

In this section we consider the case of a correlated readout noise for the covariance matrices computation. The correlated readout noise is assumed to be Gaussian distributed with a $(1/f)^{\alpha}$-like spectrum (that will be explained in Section 4.4) in Fourier domain as in [170]. In the time domain this is equivalent to Gaussian distributed noise described by a correlation function, since the correlation function is the inverse Fourier transform of the spectrum, of the form: $C\left[|\delta_t|\right]$, where $\delta_t$ is the time interval between two given frames. For computing this time interval, let's suppose that we have group $G_l$, in which the notation of each frame is $j$. If we have a number of frames per group, $m$, and a number of drops $d$, the total exposure time at a frame, $j$ of the group $G_l$, is $t_{exp} = ((l-1)*(m+d)+j)*t_{frame}$, where $t_{frame}$ is the frame integration time. If now we have a group $G_k$, with the frames denoted as $i$, the total exposure time for a frame of this group is $t_{exp} = ((k-1)*(m+d)+i)*t_{frame}$. Hence, the time interval for two frames, $j$ and $i$, belonging to the groups $l$ and $k$, respectively, would be: $\delta_t = [(l-k)*(m+d)+(j-i)]*t_{frame}$. Therefore, the covariance for the correlated readout noise can be written as:

$$< \delta\rho_i^k \delta\rho_j^l >= C[|(l-k)*(m+d)+(j-i)|*t_{frame}]. \tag{4.31}$$

Using equations 4.21 and 4.31 the group noise covariance matrix reads (see Appendix A for details):

$$\begin{aligned} C_{kk} = {} & (k-1)\,\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)(2m+1)}{6m} \\ & + \frac{1}{m^2}\left[mC(0) + 2\sum_{i=1}^{m-1}(m-i)C(i\,t_{\text{frame}})\right], \end{aligned} \tag{4.32}$$

for the diagonal terms and

$$\begin{aligned} C_{kl} = {} & (k-1)\,\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)}{2} \\ & + \frac{1}{m}\,C\left((l-k)(m+d)\;t_{\text{frame}}\right) \\ & + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C\left(|(l-k)(m+d)+(j-i)|\,t_{\text{frame}}\right), \end{aligned} \tag{4.33}$$

for the off-diagonal ones. By constrast to the white noise case described by equations 4.27, we find in the latter expression a contribution from the readout noise.

We also derive the group difference covariance matrix (see Appendix A for more details). The diagonal terms are given by:

$$\begin{aligned} D_{kk} = \mathcal{D} &+ \frac{(m-1)(2m-1)}{3m} f \\ &+ \frac{2}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} C\left(|j-i|\, t_{\text{frame}}\right) \\ &- \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} C\left(|j-i-m-d|\, t_{\text{frame}}\right) \\ &- \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} C\left(|j-i+m+d|\, t_{\text{frame}}\right), \end{aligned} \tag{4.34}$$

and the off-diagonal terms are:

$$\begin{aligned} D_{kl} = &\left(\frac{m^2-1}{6m} f\right) \delta_{l(k+1)} \\ &+ \frac{2}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} C\left(|(l-k)(m+d)+(j-i)|\, t_{\text{frame}}\right) \\ &- \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} C\left(|(l-k-1)(m+d)+(j-i)|\, t_{\text{frame}}\right) \\ &- \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} C\left(|(l-k+1)(m+d)+(j-i)|\, t_{\text{frame}}\right), \end{aligned} \tag{4.35}$$

for $k < l$. By contrast to the white noise readout noise case equations 4.30, we observe that the contribution of the readout noise to the group difference covariance is not constant in the diagonal terms and it adds extra correlation in the off-diagonal ones. In the diagonal case, comparing with the equation 4.30, the first two terms of the equation remain the same since it is the Poisson noise contribution. The main difference is the last terms related to the readout noise. As we can see, the function correlation replace the constant term proportional to $\sigma_R$ which comes from the white noise consideration. For the off-diagonal the Poisson noise only affects to the consecutive groups in the white noise case, whereas the correlation related to the readout noise can affect to frames belonging to non-consecutive groups. Now the covariance matrices are computed, the next step is to characterize the previously mentioned $(1/f)^{\alpha}$-like readout noise.

## 4.4 Characterization of the readout noise of the NISP detectors

In this section we use ground calibration data to characterize the readout noise of the NISP detectors in terms of correlated $(1/f)^{\alpha}$-like noise.

### 4.4.1 Readout noise measurements

The readout noise of infrared detectors can be characterized from long exposure ramps in dark conditions. Here, we use dark test data obtained during the *Euclid* NISP detector characterization performed at the CPPM laboratory. We focus on one of the sixteen NISP H2RG detectors, which was cooled down to a nominal temperature of 85 $K$. The testing facility was designed to achieve best possible dark conditions and special care was taken to achieve expected in-flight readout noise.

For proper dark measurement, long integration UTR ramps were acquired, typically, ramps of 8000 frames with a total exposure time of 3.21 hours corresponding to a frame exposure time of $t_f = 1.445\ s$. For each frame and for each of the 2040x2040 photosensitive pixels we use the reference pixels to remove correlations in the readout noise induced by background variations [173, 174]. First, we remove the first and last 4 pixels in a group of 9 lines. Second, we remove the average of the first and last 64x4 pixels contained in each channel. Furthermore, we compute the dark for each ramp using the Fowler-M algorithm [167], for which the slope of the ramp (in this case the dark contribution) is computed from the difference of the average of blocks of frames at the end and the beginning of the ramp, as explained in Section 4.1.3. In our case we have considered blocks of 32 frames to reduce the uncertainties in the dark measurements. Every ramp is corrected for the dark by subtracting the median dark value of all of the pixels in a given detector. For the data used in this section the median dark for all pixels in the array was about $0.006 \pm 0.002$ $e^-/s$.

The left panel of Figure 4.5 shows the measured raw data for one of these ramps for one of the inner pixels in the array after correcting for the reference pixels (raw data, red line) and after subtraction of the dark contribution (dark corrected, blue line). As the dark is very low, the contribution of photon noise is negligible for the ramp. Then, after dark subtraction, we are left with the contribution of the readout noise. We have used a conversion gain factor of $f_e = 0.5\ ADU/e^-$. We can observe in the figure that the readout noise is not fully white. This can be better seen in the right panel of the figure, where we show the power spectrum of the dark corrected data as a function of the time frequency in Hz.

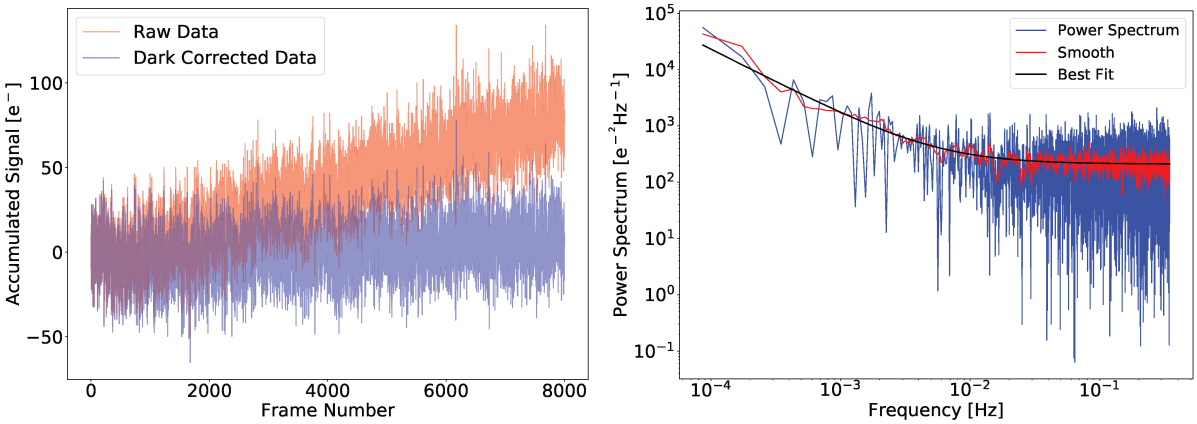


Figure 4.5 – Left panel: Accumulated signal containing both readout and photon noise obtained from a ramp before (red) and after (blue) dark and pixel reference correction for an inner pixel of the NISP array. Right panel: time power spectrum (blue) for the left panel corrected data (blue dots) and best-fit $(1/f)^{\alpha}$-like model (black). The red line represents the smoothed version of the power spectrum.

### 4.4.2 Readout noise modelling and fitting

The readout noise power spectrum seems to show a $(1/f)^{\alpha}$-like spectrum with an excess of power at low frequencies. Similar patterns are found for all other pixels in the array. Thus, to characterize the correlated readout noise in the NISP detectors we assume that its power spectrum is given by

$$P(f) = \frac{\sigma^2}{2}\,(1 + (f/f_{\rm knee})^{\alpha}), \tag{4.36}$$

with $f$ the time frequency and $\sigma$, $f_{\rm knee}$ and $\alpha$ the parameters of the model. The $\sigma$ parameter gives us information about the flat part of the power spectrum, which corresponds to the white noise contribution. $\alpha$ and $f_{knee}$ inform us about the correlated noise [2] contribution. Notice that for $\alpha = 0$ this model converges to a flat power spectrum and thus to a white noise spectrum.

For each pixel in the array we fit the readout noise power spectrum to this $(1/f)^{\alpha}$-like model. We use the python *lmfit* module [175], which gives the best-fit parameters and their uncertainties. Uncertainties on the data power spectrum are computed assuming Gaussian noise: $\sigma_{P(f)} \propto P(f)$. In practice, the fit is performed in two steps. First, we estimate the uncertainties in the power spectrum from a smoothed version of the noise power spectrum (see red line in the right panel of Figure 4.5) and compute the best-fit parameters for the $(1/f)^{\alpha}$-like model. Then, we use these first estimates of best-fit parameters to estimate the uncertainties in the power spectrum and perform a second fit to the noise power spectrum. The best-fit parameters obtained from this

2. Also called pink noise in the case of a $(1/f)^{\alpha}$-like correlated noise.

second fit are stored for further analysis. Using Monte Carlo simulations, we have observed that this two-step procedure leads to non biased estimates of the best-fit parameters for the$(1/f)^{\alpha}$-like model. In Figure 4.5 we show the best-fit $(1/f)^{\alpha}$-like model (black line) to the noise power spectrum (blue line) obtained from the second fit. The best fit-parameters and their uncertainties for this pixel are $\sigma = 20.50 \pm 0.23\ e^{-}/\sqrt{Hz}$, $f_{knee} = 0.0055 \pm 0.0008\ Hz$ and $\alpha = 1.17 \pm 0.15$. We observe in the figure that the best-fit model is consistent with the data with a reduced $\chi^2$ of 1.57.

### 4.4.3 Readout noise properties

We present in the left column of Figure 4.6 four maps representing the best-fit parameters and the reduced chi-square, $\chi^2/N_{d.o.f.}$ values for all the $2040x2040$ photosensitive pixels for one of the ramps of one of the tested detectors. The white dots in the maps correspond to either hot pixels (pixels that are saturated) or pixels for which we obtain a bad fit to the data. These pixels represent less than 0.1% of the total pixels and are uniformly distributed in the maps. We can observe in the maps vertical bands which are related to the 32 readout channels in the detector array, for which we expect some correlations in the noise properties. We can also isolate some particular regions as the one in the $f_{\rm knee}$ map for pixels around (2000,1400), which are also found when computing other characteristic quantities of the detectors as for instance the CDS noise, computed as the standard deviation of the CDS signal, explained in Section 4.1.2. They mainly correspond to manufacturing defects.

The 1D distributions of the best-fit parameters and the $\chi^2/N_{d.o.f.}$ are shown in the right panels of Figure 4.6 excluding hot and bad-fit pixels. We show in the figure four ramps of the same detector for which we find consistent results. We observe that the distributions for the three parameters are skewed towards large values. We find that the median values for the best-fit parameters are $\sigma = 19.70^{+1.11}_{-0.78}\ e^{-}/\sqrt{Hz}$, $f_{knee} = 0.0052^{+0.0018}_{-0.0013}\ Hz$ and $\alpha = 1.24^{+0.26}_{-0.21}$. We derive the uncertainties from the 15.8th ($\equiv -1\sigma$) and 84.13th ($\equiv +1\sigma$) percentiles of the distribution.

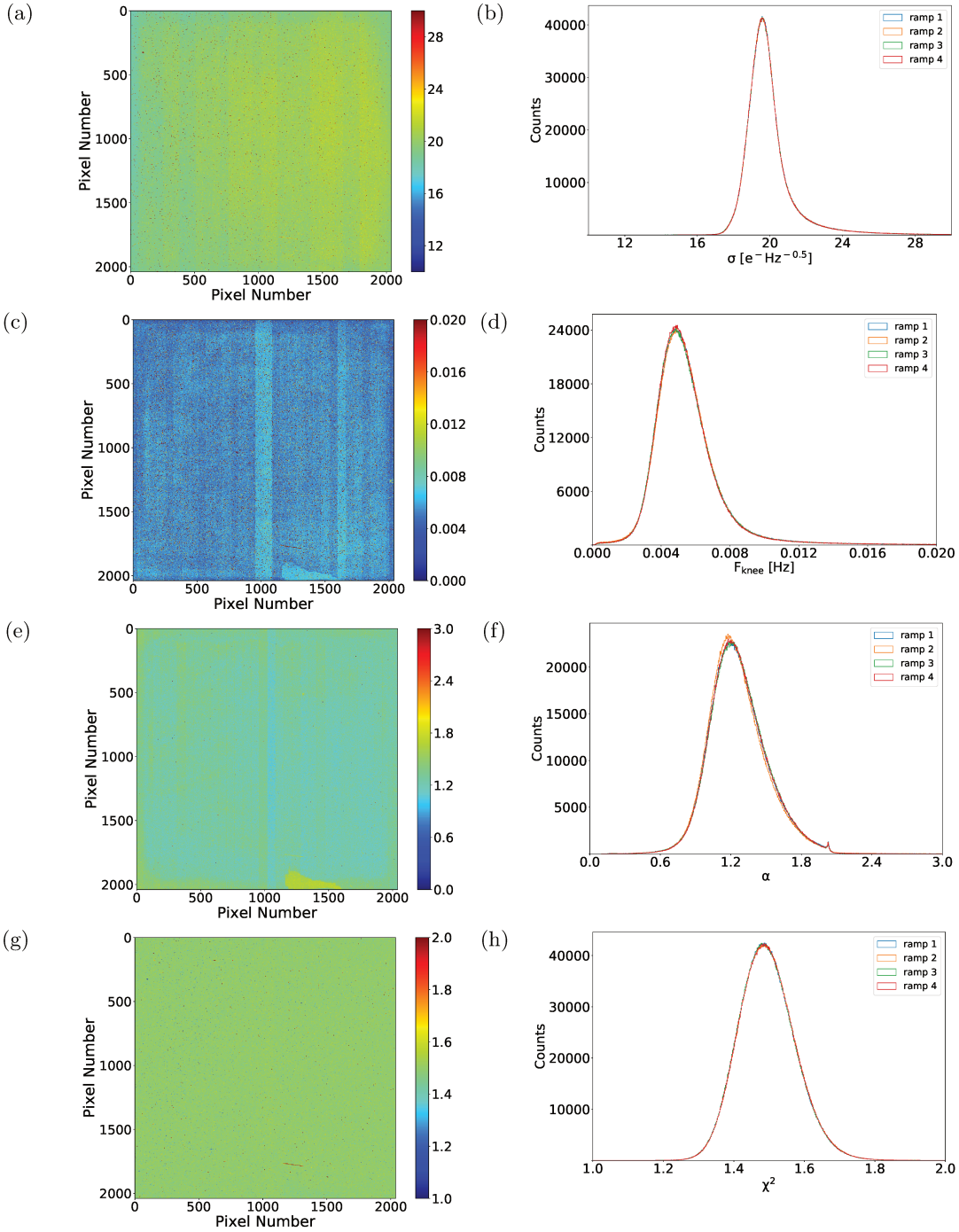


Figure 4.6 – $(1/f)^{\alpha}$-like model best-fit parameters, $\sigma$,$f_{knee}$ and $\alpha$ for the photosensitive pixels of one the NISP detectors tested. Panels (a),(c),(e) and (g) are maps representing the best-fit values for the three parameters and the $\chi^2/Nd.o.f$ for the $2040x2040$ photosensitive pixels of the detector for a single ramp. Panels (b),(d), (f) and (h) show the 1D distribution for the same best-fit parameters of 4 ramps of the same detector, and the $\chi^2/Nd.o.f$ , respectively. See main text for details.

## 4.5 Covariance Matrices for the NISP Instrument

### 4.5.1 Verification via simulations

In order to validate our analytical expressions for the group and group difference covariance matrices in the case of correlated readout noise we have performed Monte Carlo simulations. We have generated a large number of realizations of fake NISP readout noise using the $(1/f)^{\alpha}$-like model discussed in Section 4.4. The correlated readout noise simulations are obtained via three steps: 1) we produce realizations of Gaussian white noise in real space, 2) we take the Fourier transform of those and multiply each fourier component by the square root of the value of the power spectrum model at the same frequency, and 3) we compute the inverse Fourier transform of the modified Fourier components of the readout noise simulation. From these simulations of readout noise we have constructed fake NISP ramps by adding a cumulative flux contribution as well as the corresponding photon noise assuming a Poisson distribution. We assume $MACC(15, 16, 11)$ readout mode as expected for the *Euclid* spectrospy in-flight operations, also called, NISP-S mode. As an example, we present in Figure 4.7 the group difference covariance matrix as obtained from equations 4.34 and 4.35 (left panel), and from Monte Carlo simulations (right panel) for the values of $\sigma, f_{\rm knee}, \alpha$ found in Section 4.4 for the NISP detector data. The incident flux is set to $1e^{-}/s$. We observe very good agreement between the two estimates. We have repeated this comparison for various values of the parameters $\sigma$ , $f_{\rm knee}$ and $\alpha$, and for different input fluxes, and obtained the same results. We therefore validate our analytical expressions.

### 4.5.2 White and Correlated readout noise covariance matrices

In this chapter, we are interested in studying how using a white noise approximation in the case of a correlated readout noise can impact the on-board estimation of the total flux measured by the *Euclid* detectors. Therefore, it is interesting to compare the covariance matrix one would obtain for the same correlated input noise in the white and correlated readout noise approximations discussed in Section 4.3.2.

The correlated readout noise is obtained via Monte Carlo simulations. We generate mock timelines using the $(1/f)^{\alpha}$-like model discussed above with the set of averaged best-fit parameters presented in Section 4.4. The covariance matrix for the correlated noise approximation is computed as described in Section 4.3.2. For the white noise approximation we start by computing the effective rms of the readout noise. In practice we deduce it from the CDS noise (Section 4.1.2) estimated from the simulated timelines of correlated readout noise and impose $\sigma_{white} = \frac{CDS}{\sqrt{2}}$. The covariance matrix in the white noise approximation is computed following the theoretical equations 4.30. In terms of the signal contribution we have considered two extreme cases: 1)

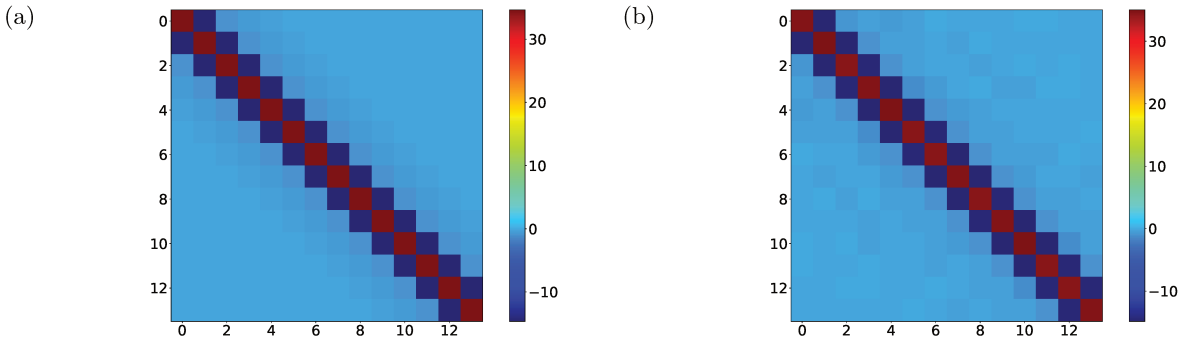


Figure 4.7 – Group difference covariance matrix as obtained from equations 4.34 and 4.35 and from Monte Carlo simulations. See text for details.

dark conditions and 2) sky background on nominal *Euclid* flight operations. For the dark conditions we assume an incident flux of $0.001e^-/s$, while for the sky background we consider $2e^-/s$.

The main results of this analysis are presented in Figure 4.8, where we represent the group difference covariance matrix for the white (left) and correlated (right) readout noise approximations. In the top and bottom panels of the figure we show the covariance matrices for the dark conditions and nominal sky background cases, respectively. For all the covariance matrices displayed we observe strong correlation between adjacent group differences as one would expect for non-destructive exposures. For the dark conditions we observe important differences between the white and correlated readout noise approximations. This is due to the fact that the readout noise dominates with respect to the photon noise. We mainly find that the difference between the diagonal and adjacent terms is increased in the correlated readout noise approximation with respect to the white noise one. Furthermore, other off-diagonal terms are not strictly zero for the correlated approximation by contrast to the white noise one. These differences would increase in the case of either steeper power spectrum or a larger $f_{\rm knee}$ frequency. For the nominal background conditions we expect the photon noise contribution to be dominant and therefore we find that the differences between the white and correlated readout noise approximations are smaller. As before, these differences would depend very much on the slope of the readout noise power spectrum and on its $f_{\rm knee}$ frequency. Therefore, we think that an accurate characterisation of these parameters will be needed during ground calibration and in-flight.

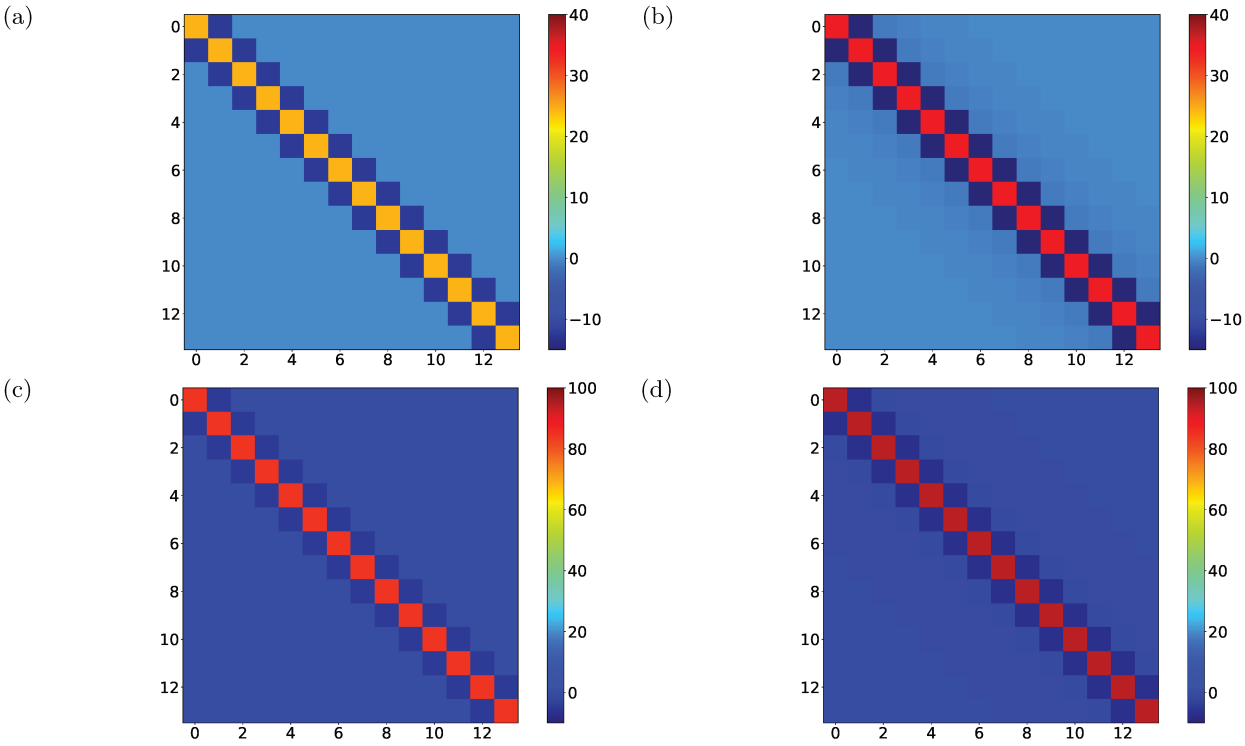


Figure 4.8 – Group difference covariance matrices in the case of the white (left) and correlated readout (right) noise approximations. The top and bottom panels correspond to an incident flux of $0.001e^-/s$ (dark conditions) and of $2e^-/s$ (nominal sky background), respectively. See main text for details

## 4.6 *Euclid* On-board flux estimation

In this section we study possible bias in the flux measurements induced by departures from the white noise approximation in the linear fit.

As discussed before during *Euclid* flight operations the maximum likelihood white noise approximation will be used to estimate the sky flux in each of the NISP array pixels and only the sky image will be transferred to Earth (the best fit ramp slope for each pixel). In the presence of correlated noise we expect the estimate of the flux to be biased. To evaluate this bias we have constructed mock simulations of the sky emission by assuming a constant sky signal and adding realistic realisations of the readout noise in the NISP detectors. For the latter we use the $(1/f)^{\alpha}$-like model and best-fit parameters discussed in Section 4.4, and the simulation procedure described above. For the sky signal we explore the range $0.1e^-/s$ to $100e^-/s$, i.e., from low dark values to bright objects.

For each value of the sky signal we construct 10000 mock ramps. For each of the simulated

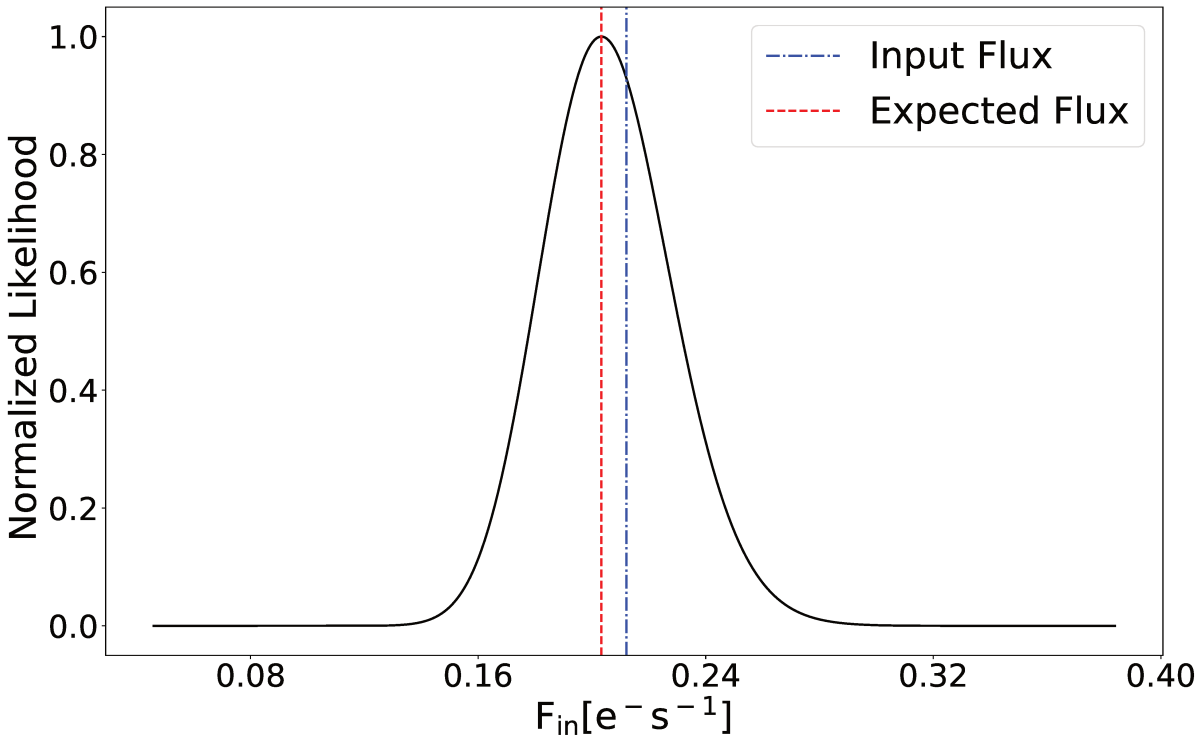


Figure 4.9 – Normalized likelihood for one simulation. The blue vertical line represents the sky signal. The red vertical line is the flux estimation through the maximum likelihood method for the correlated pink noise case.

ramps we estimate the sky flux from the group differences using: 1) the white noise approximation flux estimator (*WNA*, hereafter) developed by [176], and 2) the maximum likelihood approach discussed in equation 4.20 assuming the noise group difference covariance matrix presented in Section 4.3.2, (*CNA*, hereafter). In practice, the Maximum Likelihood flux estimate for 1) is obtained using Equation 11 in [176]. For the *CNA* case the Maximum Likelihood flux estimate is obtained using a simple grid approach. For illustration, an example of the reconstructed likelihood function for a sky flux of 0.21 $e^-/s$ is presented as solid black curve in Figure 4.9. The best-fit value and input value are indicated as vertical dashed red and dashed blue lines, respectively. We observe that even when using the expected group difference covariance matrix for the correlated readout noise there is still a small bias in the estimate of the sky flux.

## 4.7 Bias Flux Estimation

In this section we present the bias in the flux estimation from the white noise approximation and the correlated one, for both the spectroscopic and photometric modes of the NISP

instrument, NISP-S and NISP-P, respectively.

### 4.7.1 Spectroscopic NISP Mode, NISP-S

All the previous section showed results for the $MACC(15, 16, 11)$ configuration, being this one the NISP-S mode. We present in Figure 4.10 the relative bias $\frac{F_{out}-F_{in}}{F_{in}}$ in percent for both the *CNA* (red line and dots) and the *WNA* (green line and dots) cases as a function of the background flux $F_{in}$. Uncertainties in the measured bias are given by the filled red (*CNA*) and green (*WNA*) areas as computed from the Monte Carlo simulations.

For low background flux values (below $1e^-/s$) we find that the maximum bias for *CNA* is under 1% and about four times smaller than the *WNA* one. For fluxes above $1e^-/s$ the bias in both cases are equivalent and below 0.1 %. For sky observations we expect a background flux of about $1-2e^-/s$, in the region where the bias is expected to be small. However, we have observed using the dispersion over the set of Monte Carlo simulations that the *WNA* systematically underestimates the uncertainties by a factor ranging from 2 to 5. However for *CNA* the dispersion on the simulations and the measured uncertainties are consistent.

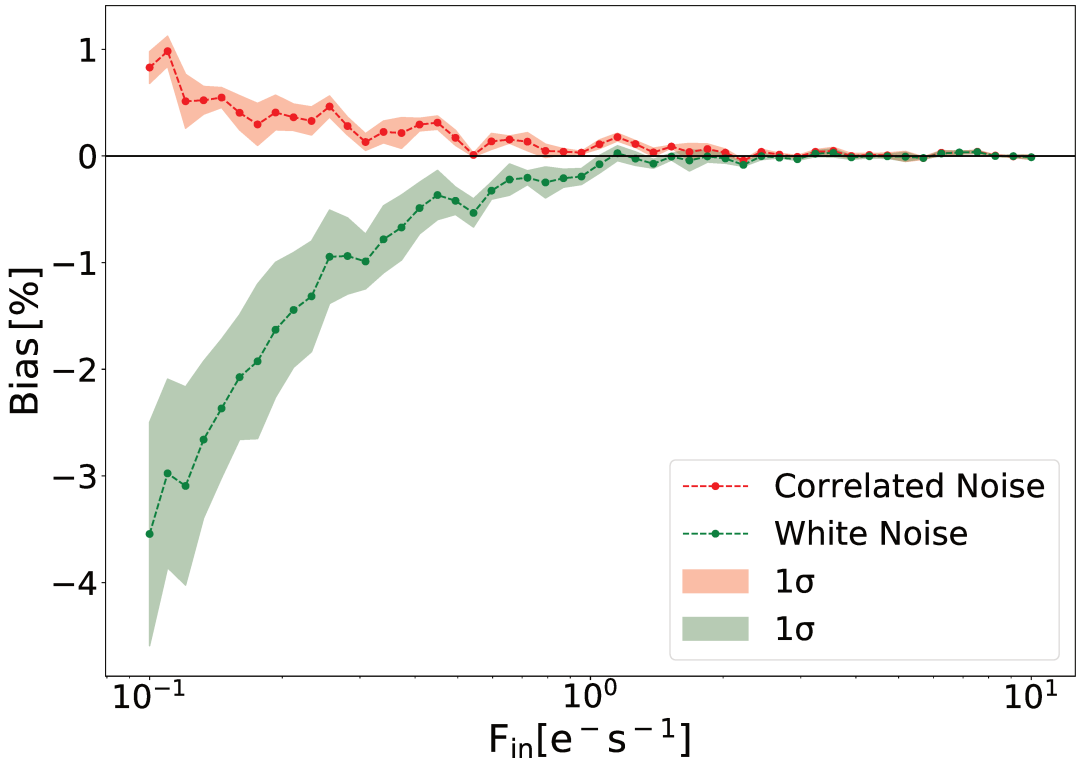


Figure 4.10 – Bias of the flux estimator in spectroscopic NIS mode, i.e., MACC(15,16,11). The green points represent the bias for the best flux estimation when considering the white noise approximation (WNA). The green area are the uncertainties coming from Monte Carlo simulations. The red points represent the bias for the best flux estimation when considering the pink noise approximation (CNA). The red area are the uncertainties from Monte Carlo simulations. The parameters for the correlated readout noise are: $\sigma = 19.70^{+1.11}_{-0.78}\ e^{-}/\sqrt{Hz}$, $f_{knee} = 0.0052^{+0.0018}_{-0.0013}\ Hz$ and $\alpha = 1.24^{+0.26}_{-0.21}$.

### 4.7.2 Photometric NISP Mode, NISP-P

For the NISP photometric mode, NISP-P, the configuration for the multiple accumulative readout mode is $MACC(4, 16, 4)$. Repeting all the process explained before in all the previous sections, we present in Figure 4.11 the relative bias $\frac{F_{out}-F_{in}}{F_{in}}$ in percent for both the *CNA* (red line and dots) and the *WNA* (green line and dots) cases as a function of the background flux $F_{in}$. Uncertainties in the measured bias are given by the filled red (*CNA*) and green (*WNA*) areas as computed from the Monte Carlo simulations.

For fluxes above $1e^-/s$ the bias in both cases are equivalent and below 0.5 %. For sky observations we expect a background flux of about $1 - 2e^-/s$, in the region where the bias is expected to be small. However, we have observed using the dispersion over the set of Monte Carlo simulations that the *WNA* systematically underestimates the uncertainties by a factor ranging from 2 to 5. However for *CNA* the dispersion on the simulations and the measured uncertainties are consistent. For low background flux values (below $1e^-/s$) we find that the maximum bias for *CNA* is not possible to compute. This can be explained with the Nyquist theorem, that says that for a high sampling rate, the relationship between the frequency and the time is given by: $t = 1/2f$. If we consider that the knee frequency for the NISP detectors readout noise is $f_{knee} = 0.0052^{+0.0018}_{-0.0013}\ Hz$, the associated time will be $t = 100s$. That means that, the correlation contribution (that is below the knee frequency as we can see in the power spectrum of the right panel of Figure 4.5) will be associated with times higher than $100s$. For the photometric bands $Y, J, H$ for which the observation times are 121s, 116s and 81s, respectively (see Figure 3.7), the correlated noise is negligible (since its associated time is close or below $t = 100s$), then we are dominated by white noise. Therefore, for the readout noise estimation in the NISP-P mode, the white noise approximation has to be considered.

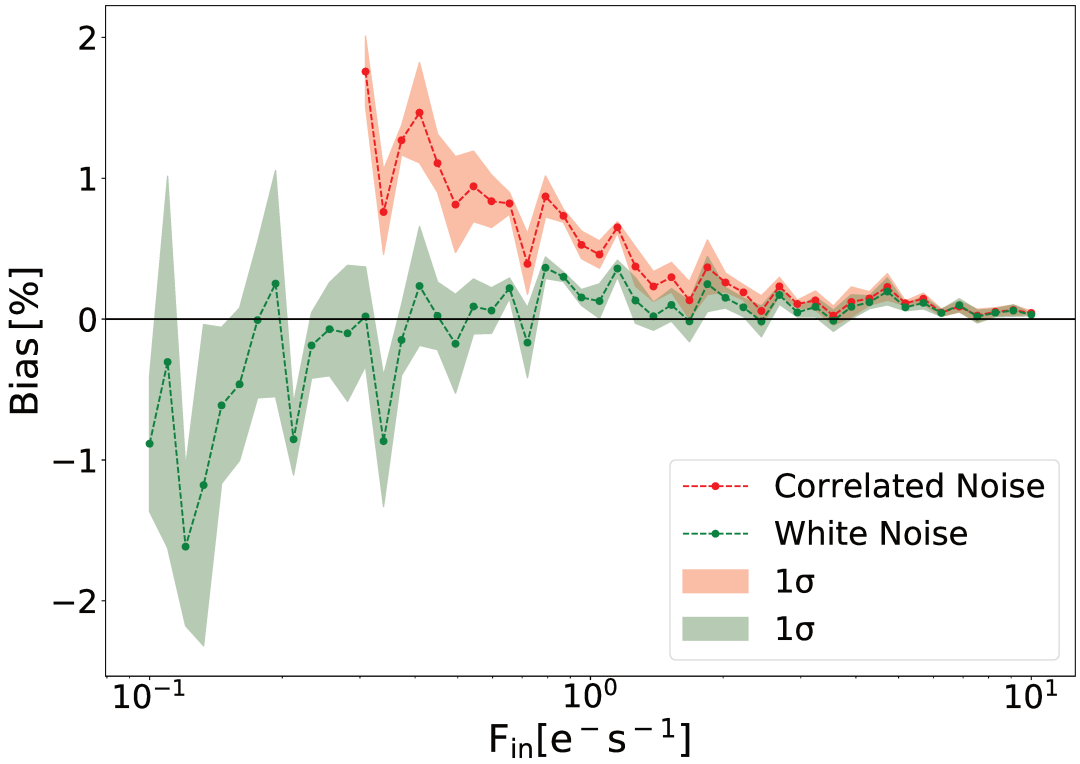


Figure 4.11 – Bias of the flux estimator in photometric NISP mode, i.e., MACC(4,16,4). The green points represent the bias for the best flux estimation when considering the white noise approximation (WNA). The green area are the uncertainties coming from Monte Carlo simulations. The red points represent the bias for the best flux estimation when considering the pink noise approximation (CNA). The red area are the uncertainties from Monte Carlo simulations. The parameters for the correlated readout noise are: $\sigma = 19.70^{+1.11}_{-0.78}$ $e^{-}/\sqrt{Hz}$, $f_{knee} = 0.0052^{+0.0018}_{-0.0013}$ $Hz$ and $\alpha = 1.24^{+0.26}_{-0.21}$.

## 4.8 Conclusion

Infrared instruments, and in particular the NISP, acquire data using the MACC readout mode, which consists of a series of non-destructive exposures averaged into groups that form a ramp. The input flux in the detectors can then be obtained from the slope of the ramp using maximum likehood estimators, which generally assume white readout noise [168, 176]. Here, we have extended these estimators to the case of correlated readout noise. Analytical expressions for the group and group difference covariance matrices are presented for the case of $(1/f)^{\alpha}$-like correlated readout noise. These have been validated via Monte Carlo simulations.

In this chapter we have studied the readout noise associated to the NISP detectors taking advantage of long exposure (few hours) performed during laboratory dark tests at the CPPM cryogenic facilities. We have found that the NISP readout noise is correlated and can be well characterized by $(1/f)^{\alpha}$-like model with a typical knee frequency of $f_{knee} = 0.0052^{+0.0018}_{-0.0013}\ Hz$ and a low frequency component with slope $\alpha = 1.24^{+0.26}_{-0.21}$. From this we conclude that the readout noise of the NISP detectors has non-negligible correlation at the typical in-flight NISP exposure time scales (574 seconds for spectroscopy mode)

Finally, we have performed Monte Carlo simulations of the in-flight expected NISP detector signal and noise, including a realistic background signal and correlated readout noise as measured on the ground calibration tests. From these simulations we have been able to estimate the expected bias in the on-board flux estimates during in-flight operations for which white readout noise is assumed. We find that for the spectroscopic mode of the NISP instrument, NISP-S, low background the flux bias can be up to four times larger than when accounting for the correlation in the readout noise. Nevertheless, this bias is negligible for typical sky background signals. Therefore, we expect no significant bias in the on-board fluxes measured by *Euclid* . On the other hand, for the photometric mode, NISP-P, the white noise approximation is the one that should be taking into account since the exposure time of the NISP-P mode corresponds to a frequency domain where the power spectrum is dominated by a white noise.

Part III

# Towards the Selection Function for the *Euclid* Survey

Chapter 5

# Optical properties from the *Euclid* Galaxy Mock catalogue



In this chapter we present the main optical cluster properties from the *Euclid* Galaxy Mock catalogue. Our goal is to reproduce and extend the analysis done in Adam et al. [134] with some variations, since our final goal is different. Our final goal is to construct analytically a cluster catalogue from observational cluster properties to test cluster finder properties as discuss in

Chapter 7.

In this chapter, first, we present how the *Euclid* Galaxy Mock and *Euclid* Cluster's Galaxy Mock catalogue are constructed. We present some cluster properties such as: cluster distribution in mass and redshift, number of cluster galaxy members as a function of cluster mass, the radial distribution of galaxies within the cluster and the distribution of galaxies as a function of their magnitude (known as Luminosity Function). Next, as proposed in Adam et al. [134], we fit analytically the galaxy density radial distribution and the luminosity function with a Navarro-Frenk-White (NFW) profile and a Schechter model, respectively.

## 5.1 *Euclid* Mock Catalogue

### 5.1.1 Construction of the *Euclid* Mock Catalogue

The *Euclid* Mock catalogue used for this thesis is extracted from Ascaso et al. [52]. It was based on Merson et al. [177], which combines a GALFORM [178] semi-analytical model (SAM) with N-body simulations. Ascaso et al. [52] includes a reprocess of the galaxy catalog to limit the Y,J,H photometry to magnitudes up to 24 at $5\sigma$, and re-estimate photometric redshifts using this photometry cut, for reproducing *Euclid*-like data. In the following we explain how Merson et al. [177] construct the simulations, for which Ascaso et al. [52] produced photometry and photometric redshift estimates. The Merson et al. [177] simulations are constructed from a N-body dark-matter-only simulations, called Millenium [23], for which only gravitation is accounted for the track of dark matter particles. In particular, in the Millenium simulation, there are $2160^3$ mass particles tracked from $z = 127$ until today in a cubic region of 500 $h^{-1}$Mpc size per side. The resolution of the dark mater haloes takes into account a detection threshold of 20 particles, meaning that the lowest available mass resolution for a halo is 1.7 x $10^{10} h^{-1} M_{\odot}$. The N-body simulations were populated with galaxies following a GALFORM [178] semi-analytical model (in particular the version of Bower et al. [179]). The semi-analytical approach incorporates galaxy properties to the N-body simulations based on theoretical and observational studies that allow one to parameterized analytical models. Some of these galaxy properties are: star formation and feedback, stellar population, chemical evolution, dust extinction, halo velocity rotation, halo density profile, etc. These properties are matched with observational data to improve the SAM simulations, for example: metallicities, galaxy colours, luminosity functions, gas fractions, and mass-to-light ratios.

The *Euclid* galaxy mock catalogue that we will use for this thesis has been constructed by the *Euclid* Cluster Finder Challenge (CFC) group as explained in Adam et al. [134], extracting from the original Ascaso et al. [52] galaxy mock catalogue, a portion of 300 $deg^2$. This catalogue

includes photometric redshifts, galaxy magnitudes and galaxy positions.

One main disadvantage of this approach using SAM simulations is that some physical processes that govern galaxy formation are not still well understood. Thus, the catalogue may not be realistic. However, this approach allows us to simulate large regions of the Universe with much less computational cost than simulations that include baryonic physics in the structure formation, called hydrodynamical simulations [54] (see Chapter 1 and Chapter 8).

### 5.1.2 The Cluster's Galaxy Mock Catalogue

From the galaxy catalogue, a cluster mock catalogue is created by identifying group of galaxies using the *Dhalo* algorithm defined in Jiang et al. [180]. The galaxies of a group are marked with an identifier as well as the central galaxy. Then the cluster catalogue is formed by galaxies of the same group (selected by their identifiers). The cluster position and redshift are set to those of the central galaxy. The difference between using the central galaxy and the barycenter of the cluster are commented in [134], which finds a negligible effect in cluster detection. For each of the mock cluster galaxy members the maximum and minimum right ascension and declination were computed for designing a rectangular area including all the galaxies of each cluster. Also the virial mass of the clusters, $M_{200}$, were computed using the *Dhalo* algorithm defined in [180] up to a 90% C.L., and then the cluster radius, $R_{200}$, having a maximum bias of 17% at 95 C.L. The final *Euclid* Cluster Mock catalogue is composed of photometric redshifts, galaxy magnitudes, galaxy positions, and cluster masses.

### 5.1.3 Properties of Galaxies and Galaxy Clusters in the Mocks

We present the main properties of the 300 deg$^2$ galaxy and cluster's galaxy mock catalogues (for more details see [52, 134]). The *Euclid* Cluster Galaxy Catalogue includes clusters with a mass in the range $10^{13}$ $M_\odot$ - $10^{15.6}$ $M_\odot$ up to a redshift of z = 3. Figure 5.1 shows the distribution of galaxy clusters as a function of mass and redshift. The distribution shows a peak at redshift 1.2 and $10^{13}$ $M_\odot$. At high redshift ($2 < z < 3$) clusters are still forming, thus they are not massive. When redshift decreases, the cluster masses increase, having a peak at about redshift one. Following this clusters' distribution is fundamental when reproducing the catalogue properties.

Richness of galaxy clusters serves as a primary mass proxy and it is derived from optical or NIR surveys. It is a fundamental parameter in cluster finder detection performance, at a given mass. Here, we define the richness as the number of galaxies associated with a cluster without any other requirement, unlike Adam et al. [134]. They estimate the richness as the galaxy cluster

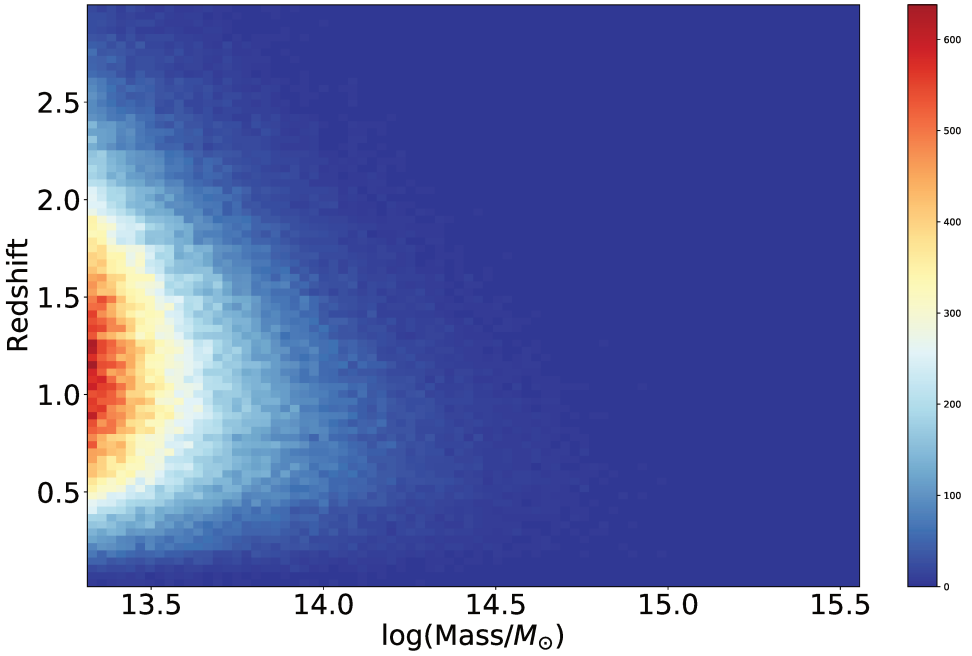


Figure 5.1 – Distribution of clusters in the *Euclid* Cluster's Galaxy Mock catalogue as a function of redshift and mass.

members for which the magnitude does not exceed $m^*_H + 2$, with $m^*_H$ a reference magnitude from a passive evolution model of a starbust galaxy, taken from the PEGASE2 library [181]. Figure 5.2 shows the richness (our definition), $\lambda$, as a function of the cluster mass, for the full redshift range (up to 3). As expected, the number of galaxies that belongs to a cluster is greater when it is more massive. Mostly the number of galaxies per cluster ranges between 10 and 100.

The next properties we want to explore are the luminosity function (LF) and the radial distribution of structures in galaxy clusters. Both are key properties that may affect the performances of cluster finders. Figure 5.3 shows the number of galaxies as a function of their magnitude in the H-band. This figure is an overall estimation of the luminosity function. Figure 5.3 shows that cluster galaxy members are mostly faint galaxies, having an important drop of number of galaxies in the bright part. This may affect cluster finders since their performance is often limited to the flux observed from the sources, i.e., it is possible that some faint galaxies are undetectable for the cluster finders. The radial distribution of galaxies depend on the cluster size, so in the next sections we will show the distribution per bins in mass and redshift.

Our main goal is to reproduce the properties of the *Euclid* Mock catalogues with analytical models, in order to construct an analytical cluster mock catalogue. The latter will be used to characterize the properties of cluster finders using an injection approach as in [138]. To

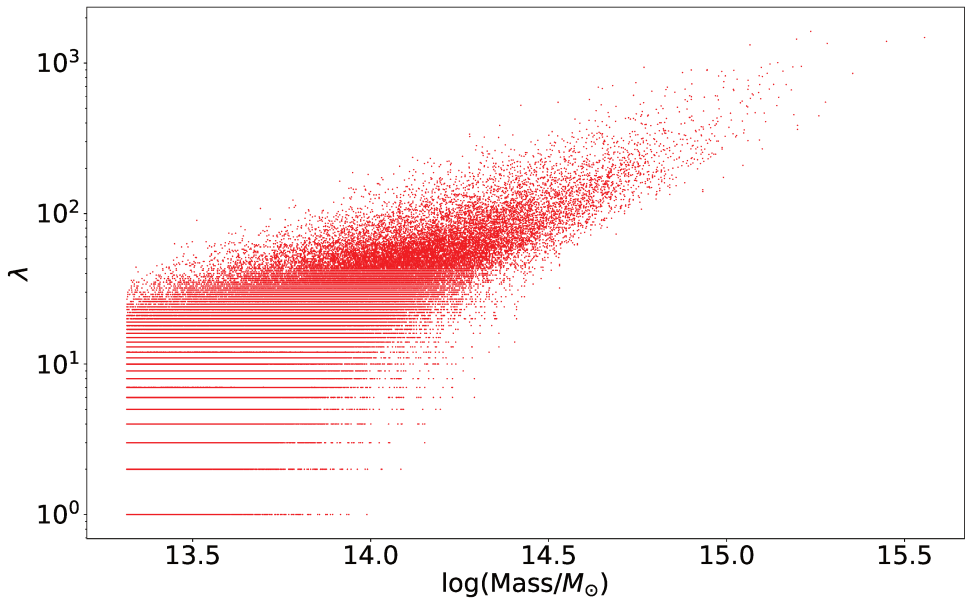


Figure 5.2 – Number of galaxies per cluster in the *Euclid* Cluster's Galaxy Mock catalogue as a function of the cluster mass.

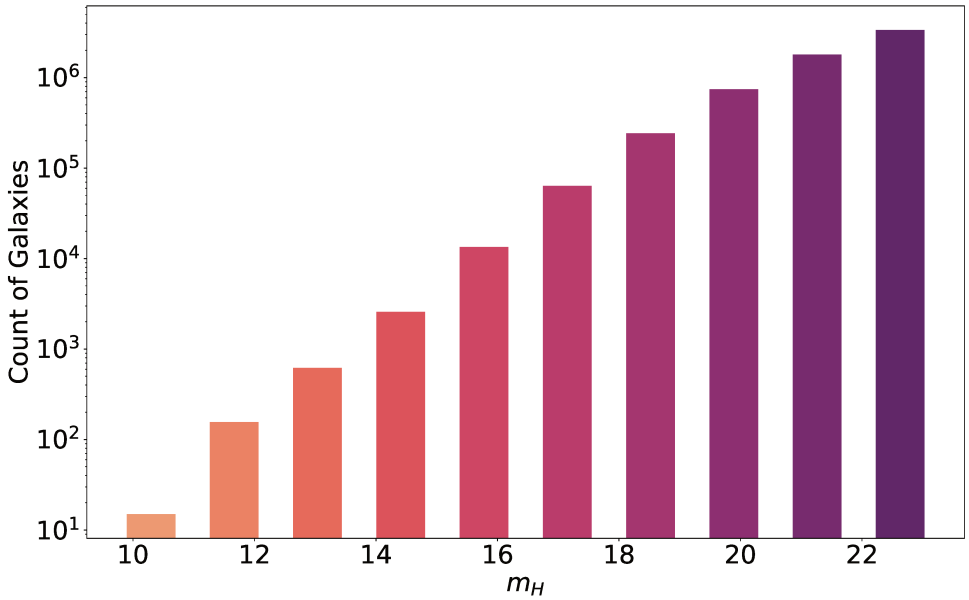


Figure 5.3 – Luminosity function in the *Euclid* Galaxy Mock catalogue as a function of cluster galaxy members magnitude.

reproduce the properties, the first thing to do is select bins in mass and redshift so that the cluster's properties are similar to each other: radius, mass, number of cluster galaxy members, distances, etc. Next, we compute the radial distribution of cluster's member galaxies and the

luminosity function. We have chosen an analytical model to fit each of these quantities, and estimate each model's parameters. Repeating this parameter estimation for several bins in mass and redshift, we will be able to construct a cluster catalogue from analytical cluster models that reproduce *Euclid* -like data.

## 5.2 Galaxy density radial distribution

The radial distribution of cluster galaxy members is a key property for cluster finders. The cluster finders used in Adam et al. [134] depend on photometric data with photometric redshift uncertainties larger than the cluster size along the line of sight. Therefore, cluster finders are more sensitive to the projected radial distribution rather than to the 3D one. For this reason we compute and fit the 2D radial distribution profiles.

To compute the galaxy density profiles, we divide the 300 deg$^2$ *Euclid* Galaxy Mock catalogue in bins of mass and redshift. We identify cluster galaxy members by their identifiers, which repeat themselves when they are associated with the same cluster. The galaxy cluster sky position and redshift are chosen to be the same as the Bright Central Galaxy (BCG). Next, we compute the angular distance, and later the radial distance, r, between each galaxy member and the BCG. The galaxy cluster size, $R_{200}$ is estimated from the clusters' redshift and mass. In the bin in mass and redshift, there are several clusters, thus to determine the radial distribution we perform a stack of each profile. These profiles depend on the cluster size, therefore, the radial distance will be normalized to $R_{200}$, i.e., $R = r/R_{200}$. The distance to the cluster center, R, is divided in equally spaced bins. For each bin, we calculate the number of galaxies for which the radial distance to the center is inside the bin, and we sum it and normalize it by the bin area and the number of clusters. We repeat the process until we are out of the cluster size and the number of cluster galaxy members drop to zero. An example of a projected stacked galaxy density profile for a bin in redshift $0 < z < 0.33$ and mass $10^{14.8}$ $M_\odot$ $< M < 10^{15}$ $M_\odot$ is shown in Figure 5.4. The black points represent the sum of the number of galaxies, N, normalized by the bin area and the number of clusters, what we will denote, from now on, as $\Sigma(R)$. The uncertainties (black error bars in the figure) are a Poisson distributed error over the number of galaxies, i.e., $\sqrt{N}$. These uncertainties are then normalized by the number of clusters and the bin area. As shown in the figure, the density of galaxies decreases at greater distance from the cluster center. However, there is a deficit of galaxies in the outskirts of the profiles, where the slope of the profile drops rapidly to zero.

### 5.2.1 Modeling and Fitting Procedure

Analytical models allow us to reproduce Mock properties with just few parameters. To fit the galaxy density profile, the model we have chosen, following Adam et al. [134], is a truncated Navarro-Frenk-White (NFW) distribution [101]. The 3D distribution of radial galaxy density can be written as

$$n(r/R_{200}) = \frac{n_0}{(cr/R_{200})(cr/R_{200}+1)^2}\mathcal{H}(r_{max}-r), \tag{5.1}$$

where c is the concentration, $n_0$ the normalization and $\mathcal{H}(r_{max}-r)$ is the Heaviside step function with $r_{max}$ the truncation radius. The latter, forces the profile to drop in the cluster outskirts, and it is not included in the original NFW distribution.

This profile is a 3D distribution, however, we want to fit to the 2D projected galaxy density. Thus we use the relationship between the projected and space number densities, $\Sigma(R)$ and $n(r)$, respectively, defined in Mamon et al. [182] as

$$\Sigma(R) = 2\int_R^\infty n(r)\frac{rdr}{\sqrt{r^2-R^2}}, \tag{5.2}$$

where r and R are the space and projected radial distances. The model parameters, $n_0$, c and $r_{max}$ are sampled using a Monte Carlo Markov Chain (MCMC) algorithm. In particular we use the Goodman & Weare's Affine Invariant MCMC Ensemble sampling [183]. The implementation of this method in Python is called *emcee*[1] [184].

The MCMC algorithms are based on the Bayes Theorem that states

$$P(theory|data) = \frac{P(data|theory)\cdot P(theory)}{P(data)}, \tag{5.3}$$

where $P(data|theory)$ is the likelihood function i.e., the probability of the data given the model (as explained in Chapter 4), $P(theory)$ is known as the prior and it is the probability of our model, which is generally transformed into allowed parameter space and $P(data)$ is the probability of the data. Finally, $P(theory|data)$ is called the posterior, that is the probability of our model given the data, and the function that want to be sampled by the MCMC algorithm. The method is applied as follows:

1. We define the prior as an uniform distribution for: $10 < n_0 < 5$ x $10^4$, $0 < c < 50$ and $0.1 < r_{max} < 10$, and zero otherwise.

1. An example of how this method is applied can be found in https://emcee.readthedocs.io/en/stable/tutorials/line/

2. Our likelihood function is assumed to be Gaussian and defined as

$$\ln P(data|n_0, c, r_{max}) = -\frac{1}{2}\sum \frac{(data - \Sigma(R))^2}{\sigma^2_{data}} + \log \sigma^2_{data}, \tag{5.4}$$

   where data and $\sigma_{data}$ are the *Euclid* projected galaxy density distribution data and its uncertainties (black points with error bars of Figure 5.4). $\Sigma(R)$ is the projected NFW model, as given by Equation 5.2.

3. The data probability, P(data), is set to one. No particular weighting is applied.

4. The posterior distribution is sampled by initializing the parameters (or walkers) in a small Gaussian around the maximum likelihood. Next, we run 200 steps of MCMC for 200 walkers. Combining all the walkers we obtain chains of 40000 samples. The first 50 values of the chain are discarded for burn-in. After few steps the walkers starts to explore the full parameter space and the posterior distribution.

5. The one and two dimensional projections of the posterior probability distribution of the parameters for the data in Figure 5.4 are shown in the corner plot of Figure 5.5. This is called a corner plot. This plot shows the covariances between the parameters. As we can see, the concentration and normalization parameters are degenerate, meaning that the parameters are not independent of each other. We observe a similar behaviour between $r_{max}$ with both concentration and normalization. Thus, none of the three parameters are independent. The degeneracy between parameters explains why there are more than one peak on the histograms, i.e., there are more than one combination of parameters that can fit our data. For the bin we are considering, the concentration varies between 9 and 12, which is in agreement with Adam et al. [134].

6. Using the full parameters' probability distributions, we calculate the best-fit value with its uncertainties based on the 16th, 50th and 84th percentiles.

The results of the fit are shown in Figure 5.4, where the blue dashed line represents the best-fit value and the red and green shaded areas the $1\sigma$ and $2\sigma$ uncertainties. In this case the NFW model is a good fit to the data.

We repeat the process of computing the 2D galaxy density profiles and fitting them by a truncated NFW, for the full redshift and mass range: $10^{13} M_\odot < M < 10^{15.6} M_\odot$ and $0.0008 < z < 3$. Some examples of 2D galaxy density profiles with their best-fit are shown in Figure 5.6 and their respective corner plots in Figure 5.7. See the caption of the figures for more details. Overall, we observe that for clusters with masses below $10^{14} M_\odot$ (upper left, and bottom right panels of the figures) the truncated NFW does not fit properly the data, as well as for redshifts above $z > 2$. We will consider clusters with $M > 10^{14} M_\odot$ and $z < 2$ for the results in the next chapters. In the corner plots we observe the same dependency behaviour between the parameters

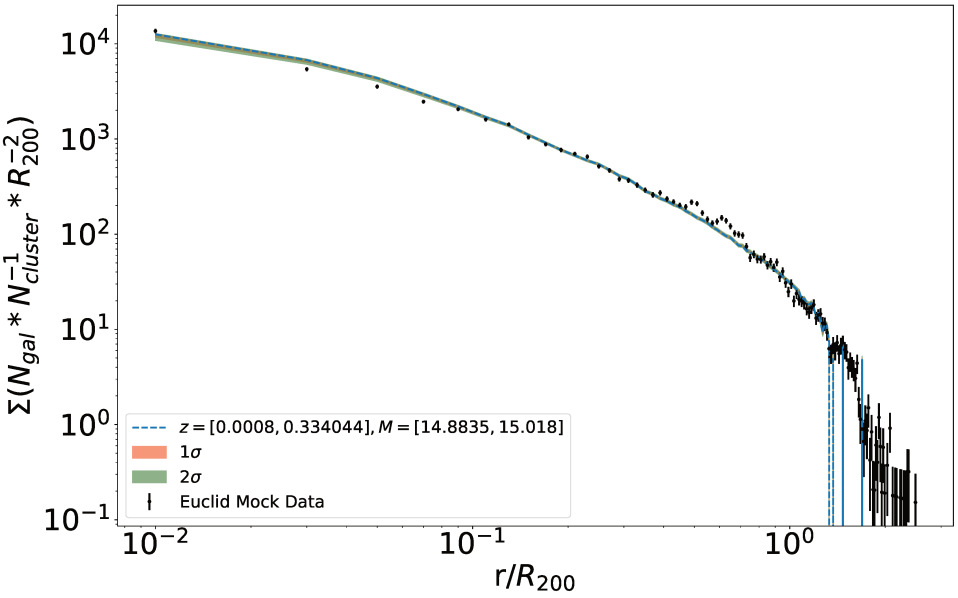


Figure 5.4 – 2D galaxy density radial distribution fitted to a truncated NFW profile. The bin in redshift is $0 < z < 0.33$ and in Mass $10^{14.8} M_\odot < M < 10^{15} M_\odot$. The black points with the error bars are the number of galaxies per cluster per bin area with a Poisson distributed uncertainty over the number of galaxies. The blue dashed line and the red and green shaded areas are the best-fit and $1\sigma$ and $2\sigma$ spread, respectively.

as shown in Figure 5.5. The concentration parameter is small at high mass. When mass increases but redshift decreases, we observe, comparing the bottom left panel of Figure 5.6 with Figure 5.4, that the concentration decreases. In contrast, when redshift increases and mass decreases the concentration increases, as we observe when comparing the bottom right panel of Figure 5.6 with Figure 5.4. These results tell us that there is a dependency of the concentration both with redshift and mass.

### 5.2.2 Cluster Galaxy Density Variation with Mass and Redshift

We repeat the fit for the different bins in mass and redshift for covering the full *Euclid* range. Our goal is to be able to reproduce the *Euclid* Cluster's Galaxy Mock Catalogue properties, and construct a Mock Catalogue from parametric models. For this, we record the best-fit value of the model parameters, $n_0$ and c, for each profile fit at every bin in mass and redshift. The drop in the galaxy density in the cluster outskirts is a simulation issue and it will not be observed with Euclid, thus we decide not to take into account the $r_{max}$ parameter when reproducing analytically the profiles. Figure 5.8 shows the best fit amplitude parameter $n_0$, color coded, as a function of mass and redshift. The colorbar represents the actual value of the parameter. We observe no clear evolution with redshift nor the mass. There are some bins, specially at low mass

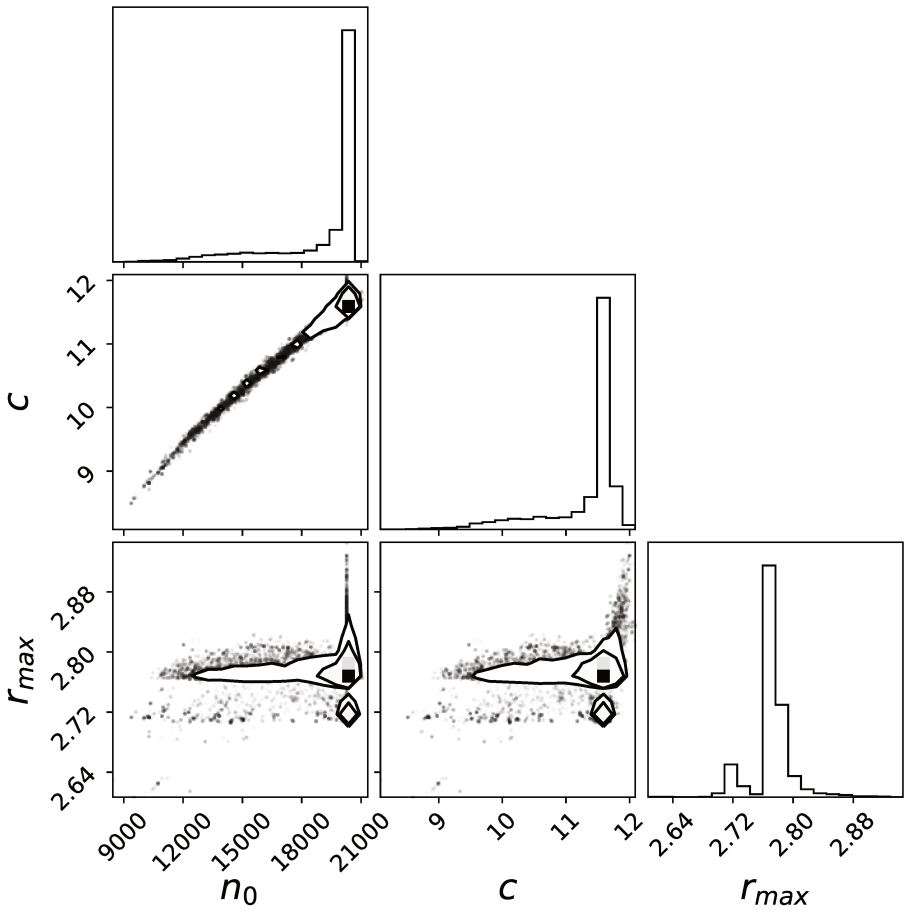


Figure 5.5 – 1D and 2D dimensional projections of the posterior probability function for the truncated NFW model parameters. The bin in redshift is $0 < z < 0.33$ and $10^{14.8}\mathrm{M}_{\odot} < \mathrm{M} < 10^{15}\mathrm{M}_{\odot}$ in mass.

or high redshift with significant difference. These bins represent extreme rare cases for which is difficult to find clusters and to have sufficient statistics. Not taking into account these cases, we observe a constant behaviour for this parameter, with values of the order of 1 x $10^4$.

Figure 5.9 shows the best fit value parameter c, color coded, as a function of mass and redshift. The colorbar represents the actual value of the parameter. There are some bins, specially at low mass or high redshift with significant difference, as for $n_0$. These bins represent extreme rare cases for which is difficult to find clusters and have sufficient statistics. Not taking into account these cases, we observe a clear evolution with redshift and mass. At higher mass values, the concentration does not depend on redshift, on agreement with Adam et al. [134]. For constant redshift, concentration decreases with mass, also in agreement with [134]. However, at low mass,

the concentration increases with redshift, not remaining constant as we observe at high mass.

## 5.3 Luminosity Function

The distribution of galaxies as a function of their magnitude, or Luminosity Function (LF), of cluster galaxy members is also a key property for cluster finders. To compute the LF, we follow a similar procedure as for the galaxy density profiles, explained in Section 5.2. We divide the 300 deg$^2$ *Euclid* Galaxy Mock catalogue in bins of mass and redshift. We identify cluster galaxy members by their identifiers, which repeat themselves when they are associated to the same cluster. The galaxy cluster sky position and redshift are chosen to be the same as the Bright Central Galaxy (BCG). Next, we compute the angular distance, and later the radial distance, r, between each galaxy member and the BCG. The galaxy cluster size, $R_{200}$ is estimated from the clusters' redshift and mass. We select galaxies within $R_{200}$. If in the bin in mass and redshift, there are several clusters, to determine the LF we perform a stack of each cluster distribution. We compute the total area defined by all the clusters in the bin, i.e., $A = \sum \pi\ R^2_{i,200}$ where i, is an index for each cluster. From the galaxy catalogue we extract the apparent magnitude in the H-band for each cluster galaxy member. The magnitude distribution is divided in equally spaced bins. For each bin, we calculate the number of galaxies with magnitude inside the bin, and we sum it and normalize it by the total area, A, and the magnitude difference in the bin, i.e., $\Delta m_H$. We repeat the process until we are out of the magnitude bin. An example of a LF for a bin in redshift $0 < z < 0.3$ and mass $10^{14.8}\ M_\odot < M < 10^{15}\ M_\odot$ is shown in Figure 5.10. The black points represent the LF, as defined above, denoted $\Phi(m)$. The uncertainties (black error bars in the figure) are a Poisson distributed error computed from the number of galaxies N, i.e., $\sqrt{N}$. These uncertainties are then normalized by the bin magnitude difference, $\Delta m_H$ and the total clusters area, A. As shown in the Figure, most of the galaxies are in the faint part of the LF, represented by high apparent magnitudes. However, the bright ones, a priori, are the easiest to detect.

### 5.3.1 Modeling and Fitting Procedure

To fit the LF, we have chosen, following Adam et al. [134], a Schechter model [93, 97], written as

$$\Phi(m) = 0.4 \ln(10) \phi^* 10^{0.4(m^* - m)(\alpha+1)} \exp\left(-10^{0.4(m^* - m)}\right) \tag{5.5}$$

with $\phi^*$ the normalization, $m^*$ the characteristic magnitude and $\alpha$ the faint-end slope. The parameters are sampled using a Monte Carlo Markov Chain (MCMC) method, as explained in Section 5.2.1. The method, in this case, is applied as follows:

1. We define the prior as an uniform distribution between: $0 < \phi^* < 100$, $10 < \mathrm{m}^* < 30$ and $-3 < \alpha < 0$. Otherwise the function is set to zero. The main difference with Adam et al. [134] is that we do not set an upper limit in magnitude to fit the model in $\mathrm{m}^* + 2$. Our main goal is different to theirs, because we want to reproduce the cluster properties analytically, while they wanted to analyze the physical cluster properties. So we need to fit the Schechter function in all the photometric range, up to $\mathrm{m} = 24$.
2. Our likelihood function is assumed to be Gaussian and defined as

$$\ln P(data|\phi^*, m^*, \alpha) = -\frac{1}{2}\sum \frac{(data - \Phi(m))^2}{\sigma^2_{data}} + \log \sigma^2_{data}, \tag{5.6}$$

   where data and $\sigma_{data}$ are the luminosity function and its uncertainties (black points with error bars of Figure 5.10) and $\Phi(m)$ is the Schechter model, as in Equation 5.5.
3. The data probability, P(data), is set to one. No particular weighting scheme is used.
4. We run 500 steps of MCMC for 500 walkers. We obtain chains of 250000 samples in total. The first 50 values of the chain are discarded for burning. After a few steps the walkers start to explore the full parameter space and the posterior distribution.
5. The one and two dimensional projections of the posterior probability distribution of the parameters are shown in Figure 5.11. We can see that the characteristic magnitude, the faint-end slope and normalization are degenerate. For the bin in mass and redshift, that we are considering, the characteristic magnitude varies between 16.5 and 17, which is in agreement with Adam et al. [134].
6. From the posterior probability distribution we compute the best-fit with its uncertainties based on the 16th, 50th and 84th percentiles.

The results of the fit are shown in Figure 5.10, where the blue dashed line represents the best-fit value and the red and green shaded areas the $1\sigma$ and $2\sigma$ uncertainties. The theoretical Schechter model is a good fit to the data in the faint part of the LF, but not for the bright galaxies.

We repeat the process of computing the LF profiles and fitting them by a Schechter model, for the full redshift and mass range shown in Figure 5.1: $10^{13}\mathrm{M}_\odot < \mathrm{M} < 10^{15.6}\mathrm{M}_\odot$ and $0.0008 < z < 3$. Some examples of LF profiles with their best-fit are shown in Figure 5.12 and their respective corner plots in Figure 5.13. See the caption of the figures for more details. The corner plots show the same dependency behaviour between the parameters shown in Figure 5.11. The characteristic magnitude increases with redshift. So does the normalization, $\phi^*$. The faint-end slope, varies, both increasing and decreasing, when comparing with Figure 5.10, so we observe a dependency with mass and/or redshift.

### 5.3.2 Luminosity Function Variation with Mass and Redshift

Our goal is to be able to reproduce the *Euclid* Cluster's Galaxy Mock Catalogue properties, and construct a Mock Catalogue from analytical models. For this we record the best-fit value of the model parameters, $\phi^*$, m$^*$, and $\alpha$, for each LF profile fit at every bin in mass and redshift. Figure 5.14 shows in upper left the best-fit parameter $\phi^*$, color coded, as a function of mass and redshift. The colorbar represents the value of the parameter. We observe that mostly it remains constant for values under 20. However, at medium redshifts, between 1 and 2, the normalization increases. In this region we find most of the clusters, as observed in Figure 5.1. Therefore, the total number of galaxies per bin is different and this can explain the behaviour of the parameter. On the other hand, the Schechter model does not fit the bright part of the LF for several bins in mass and redshift, as seen for example in Figure 5.10. This drop in the total number of galaxies can affect the normalization parameter probability distribution. Therefore, there is not a clear physical explanation on the behaviour of $\phi^*$ with mass and redshift.

Figure 5.14 shows in upper right plot the best-fit parameter m$^*$, color coded, as a function of mass and redshift. The colorbar represents the value of the parameter. The characteristic magnitude varies from 16 to 24. These values are shifted with respect to Adam et al. [134] (see Figure 5), where m$^*$ varies from 13 to 22. This is due to the lack of bright galaxies when performing the LF fit. However, the parameter evolution with mass and redshift is in agreement with Adam et al. [134]. It does not evolve with mass. A priori, the size of the cluster does not affect the photon flux we observe from galaxies. Therefore, the magnitude should not be affected by the cluster's mass, as observed in the Figure. On the other hand, there is an evolution with redshift. At higher redshift, greater is the apparent characteristic magnitude. If the cluster is farther, the flux we observe is lower, resulting in fainter galaxies, in other words, higher apparent magnitude.

Figure 5.14 shows in the bottom row the best fit amplitude parameter $\alpha$, color coded, as a function of mass and redshift. The colorbar represents the actual value of the parameter. The faint-end slope varies from 0 to -3.5. This is in agreement with literature [185], where for redshifts below 0.7 the LF for the XXL survey [186] the faint-end slope varies from -1 to -1.5, as our case. As observed with the characteristic magnitude, there is an evolution with redshift but not with mass.

## 5.4 Conclusion

In this chapter we have studied two properties of cluster that are fundamental for the performance of the cluster detection algorithms, the galaxy radial distribution and the luminosity function. The goal is to reproduce analytically these properties to produce a cluster catalogue

for its later injection in the Euclid Cluster's Galaxy Mock catalogue. Then, applying cluster finders on the cluster injected mock catalogue.

In the galaxy density radial distribution, we have found that the profiles are well described by a NFW model when adding a truncation radius due to the lack of galaxies in the outskirts of the clusters. The evolution of the concentration parameter with mass and redshift show that more massive clusters have less concentration. On the other hand, far clusters (high redshift) and low mass clusters present higher concentrations. These results are in agreement with Adam et al. [134]. The normalization and concentration parameters are highly degenerated. We conclude that although there are caveats some the NFW represents sufficiently well the *Euclid* Mock for our purpose.

For the luminosity function, we chose a Schechter model to fit the data, as done in Adam et al. [134] We observe, as expected, a clear evolution with the redshift, presenting fainter galaxies at higher redshift. The characteristic magnitude and faint-end slope are degenerate with the normalization. We find that the Schechter model does not fit the bright part of the luminosity function when apparent magnitudes are below 16. In other words, our analytical model present a deficit of bright galaxies with respect to the *Euclid* Mock catalogue. Thus, we have decided not to use it in Chapter 6 for simulating clusters. We will use instead the measured LF in the various bins in mass and redshift.

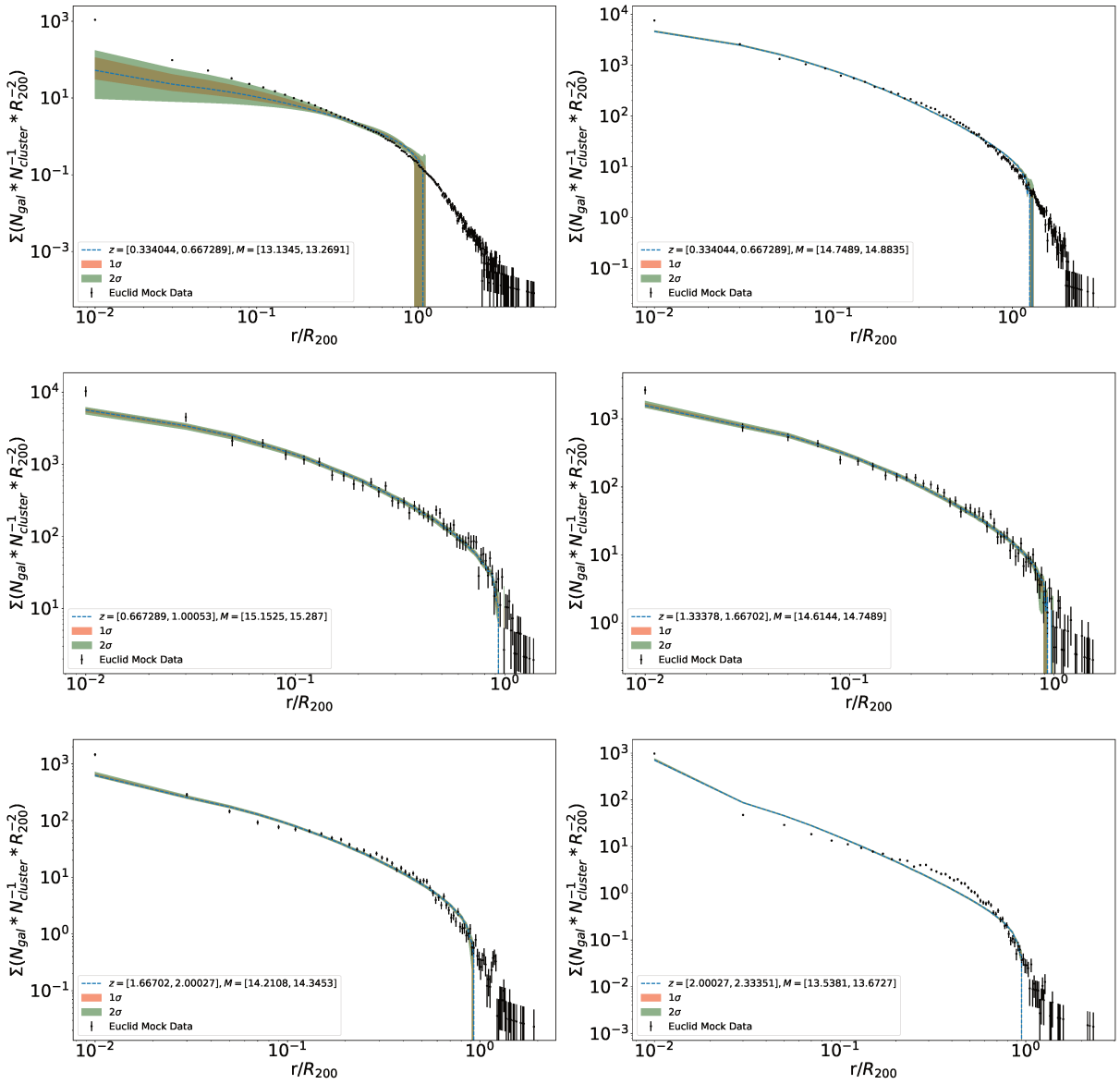


Figure 5.6 – 2D galaxy density radial distributions fitted to a truncated a NFW profile. For the upper row the bins are: in redshift is $0.3 < z < 0.7$ and in mass $10^{13.13}\mathrm{M}_{\odot} < \mathrm{M} < 10^{13.27}\mathrm{M}_{\odot}$ (left) and in redshift is $0.3 < z < 0.7$ and in mass $10^{14.74}\mathrm{M}_{\odot} < \mathrm{M} < 10^{14.88}\mathrm{M}_{\odot}$ (right). The middle row represents the bins: in redshift is $0.6 < z < 1$ and in mass $10^{15.15}\mathrm{M}_{\odot} < \mathrm{M} < 10^{15.28}\mathrm{M}_{\odot}$ (left) and in redshift is $1.3 < z < 1.6$ and in mass $10^{14.6}\mathrm{M}_{\odot} < \mathrm{M} < 10^{14.74}\mathrm{M}_{\odot}$ (right). The bottom row represents the bins: in redshift is $1.6 < z < 2$ and in mass $10^{14.2}\mathrm{M}_{\odot} < \mathrm{M} < 10^{14.34}\mathrm{M}_{\odot}$ (left) and in redshift is $2 < z < 2.3$ and in mass $10^{13.53}\mathrm{M}_{\odot} < \mathrm{M} < 10^{13.67}\mathrm{M}_{\odot}$ (right). The black points with the error bars are the number of galaxies per cluster per bin area with a Poissonian uncertainty over the number of galaxies. The blue dashed line and the red and green shaded areas are the best-fit and $1\sigma$ and $2\sigma$ spread, respectively.

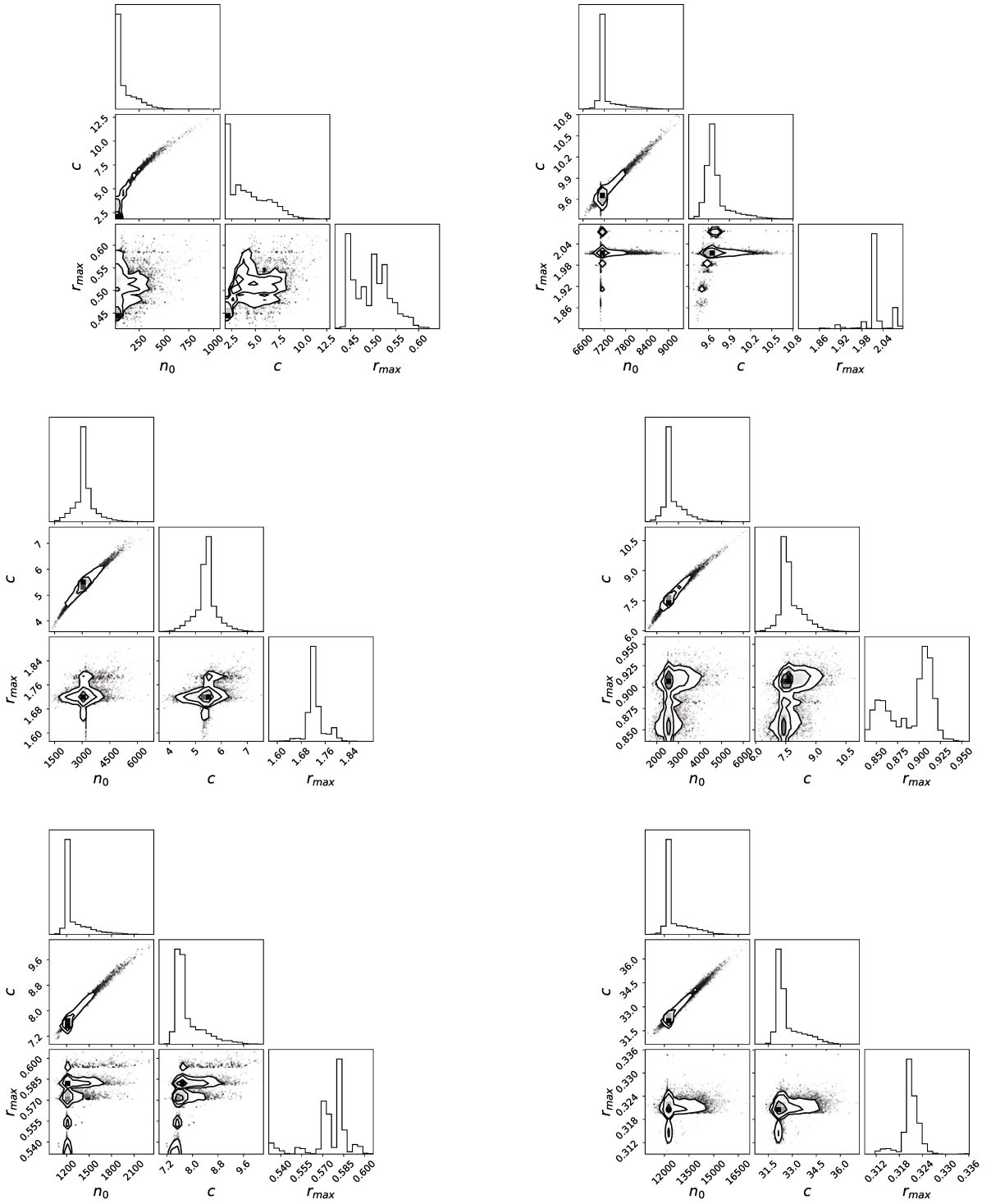


Figure 5.7 – Corner plots of the 2D galaxy density radial distributions fitted to a truncated a NFW profile. For the upper row the bins are: in redshift is $0.3 < z < 0.6$ and in mass $10^{13.13}\mathrm{M}_\odot < \mathrm{M} < 10^{13.27}\mathrm{M}_\odot$ (left) and in redshift is $0.3 < z < 0.6$ and in mass $10^{14.74}\mathrm{M}_\odot < \mathrm{M} < 10^{14.88}\mathrm{M}_\odot$ (right). The middle row represents the bins: in redshift is $0.6 < z < 1$ and in mass $10^{15.15}\mathrm{M}_\odot < \mathrm{M} < 10^{15.28}\mathrm{M}_\odot$ (left) and in redshift is $1.3 < z < 1.6$ and in mass $10^{14.6}\mathrm{M}_\odot < \mathrm{M} < 10^{14.74}\mathrm{M}_\odot$ (right). The bottom row represents the bins: in redshift is $1.6 < z < 2$ and in mass $10^{14.2}\mathrm{M}_\odot < \mathrm{M} < 10^{14.34}\mathrm{M}_\odot$ (left) and in redshift is $2 < z < 2.3$ and in mass $10^{13.53}\mathrm{M}_\odot < \mathrm{M} < 10^{13.67}\mathrm{M}_\odot$ (right).

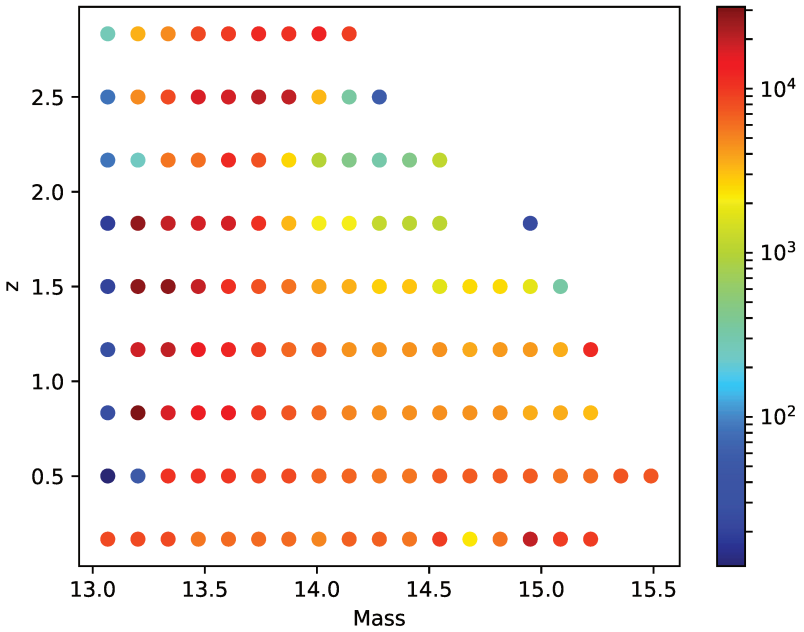


Figure 5.8 – Color coded best-fit parameter, $n_0$, as a function of mass and redshift.

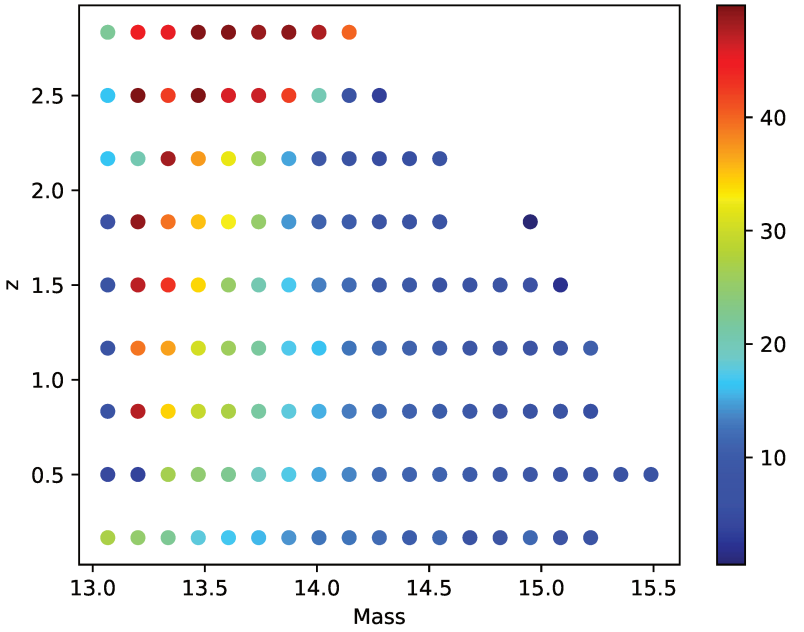


Figure 5.9 – Color coded best-fit parameter, c, as a function of mass and redshift.

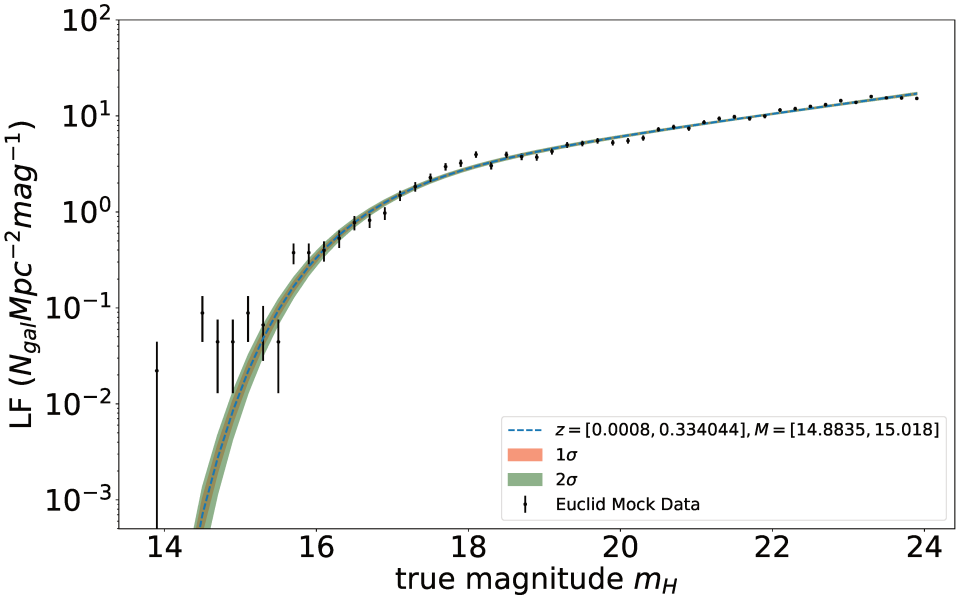


Figure 5.10 – Luminosity function fitted to a Schechter model. The bin in redshift is $0 < z < 0.33$ and in Mass $10^{14.8}\mathrm{M}_{\odot} < \mathrm{M} < 10^{15}\mathrm{M}_{\odot}$. The black points with the error bars correspond to the LF. The blue dashed line and the red and green shaded areas are the best-fit and $1\sigma$ and $2\sigma$ uncertainties, respectively.

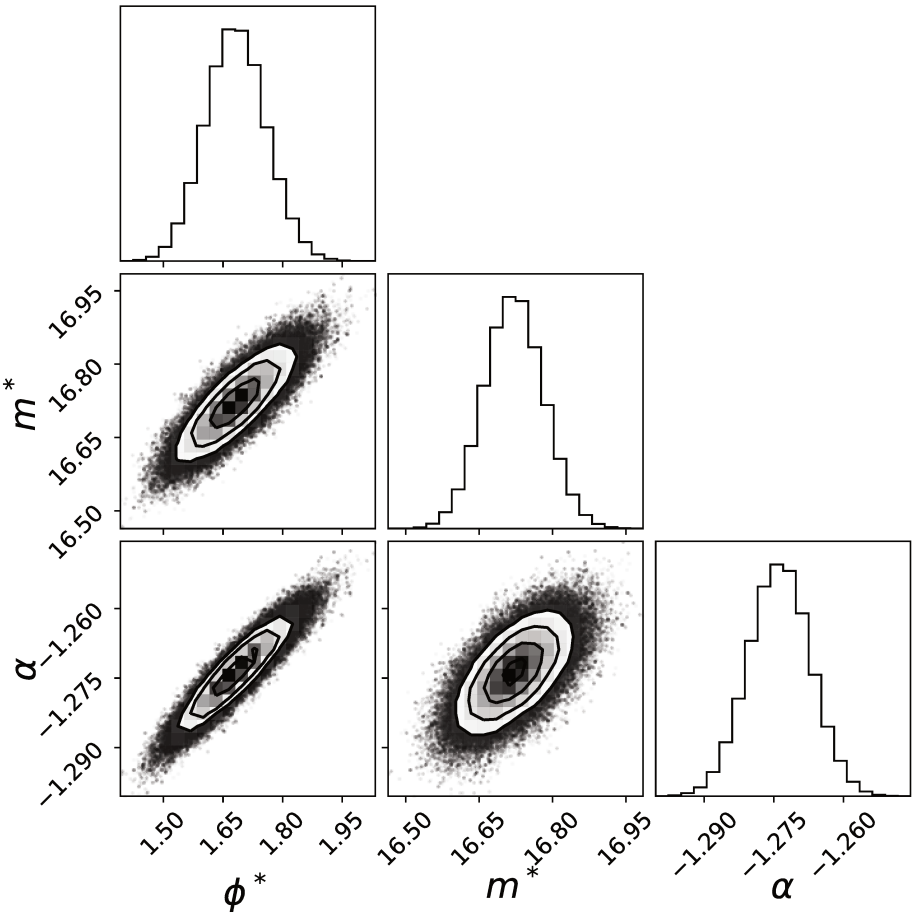


Figure 5.11 – 1D and 2D dimensional projections of the posterior probability function for the Schechter model parameters. The bin in redshift is $0 < z < 0.33$ and in Mass $10^{14.8}\mathrm{M}_\odot < \mathrm{M} < 10^{15}\mathrm{M}_\odot$.

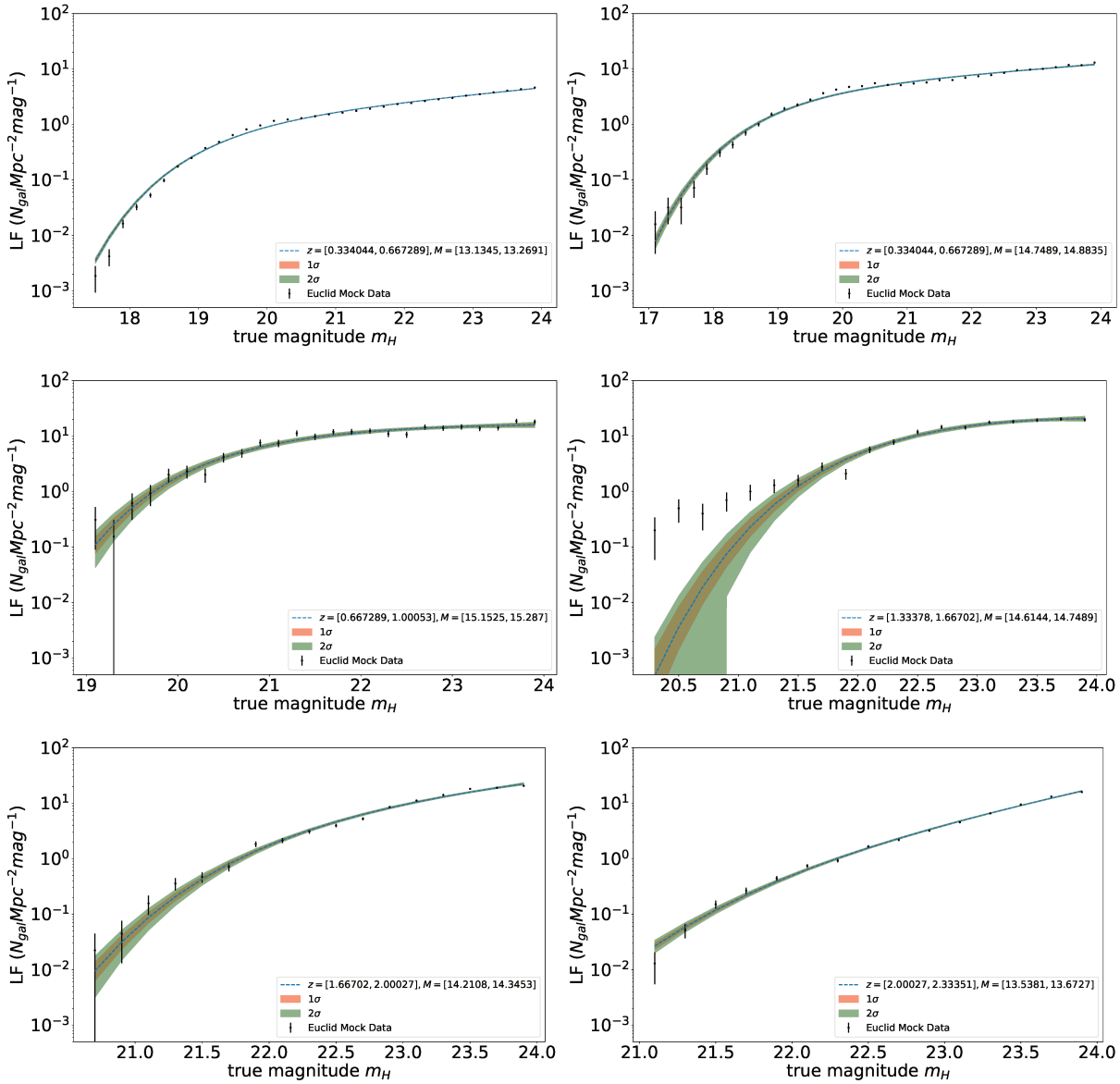


Figure 5.12 – The LF (black points) fitted to a Schechter model. For the upper row the bins are: in redshift is $0.3 < z < 0.6$ and in mass $10^{13.13}\mathrm{M}_\odot < \mathrm{M} < 10^{13.27}\mathrm{M}_\odot$ (left) and in redshift is $0.3 < z < 0.6$ and in mass $10^{14.74}\mathrm{M}_\odot < \mathrm{M} < 10^{14.88}\mathrm{M}_\odot$ (right). The middle row represents the bins: in redshift is $0.6 < z < 1$ and in mass $10^{15.15}\mathrm{M}_\odot < \mathrm{M} < 10^{15.28}\mathrm{M}_\odot$ (left) and in redshift is $1.3 < z < 1.6$ and in mass $10^{14.6}\mathrm{M}_\odot < \mathrm{M} < 10^{14.74}\mathrm{M}_\odot$ (right). The bottom row represents the bins: in redshift is $1.6 < z < 2$ and in mass $10^{14.2}\mathrm{M}_\odot < \mathrm{M} < 10^{14.34}\mathrm{M}_\odot$ (left) and in redshift is $2 < z < 2.3$ and in mass $10^{13.53}\mathrm{M}_\odot < \mathrm{M} < 10^{13.67}\mathrm{M}_\odot$ (right). The blue dashed line and the red and green shaded areas are the best-fit and $1\sigma$ and $2\sigma$ uncertainties, respectively.

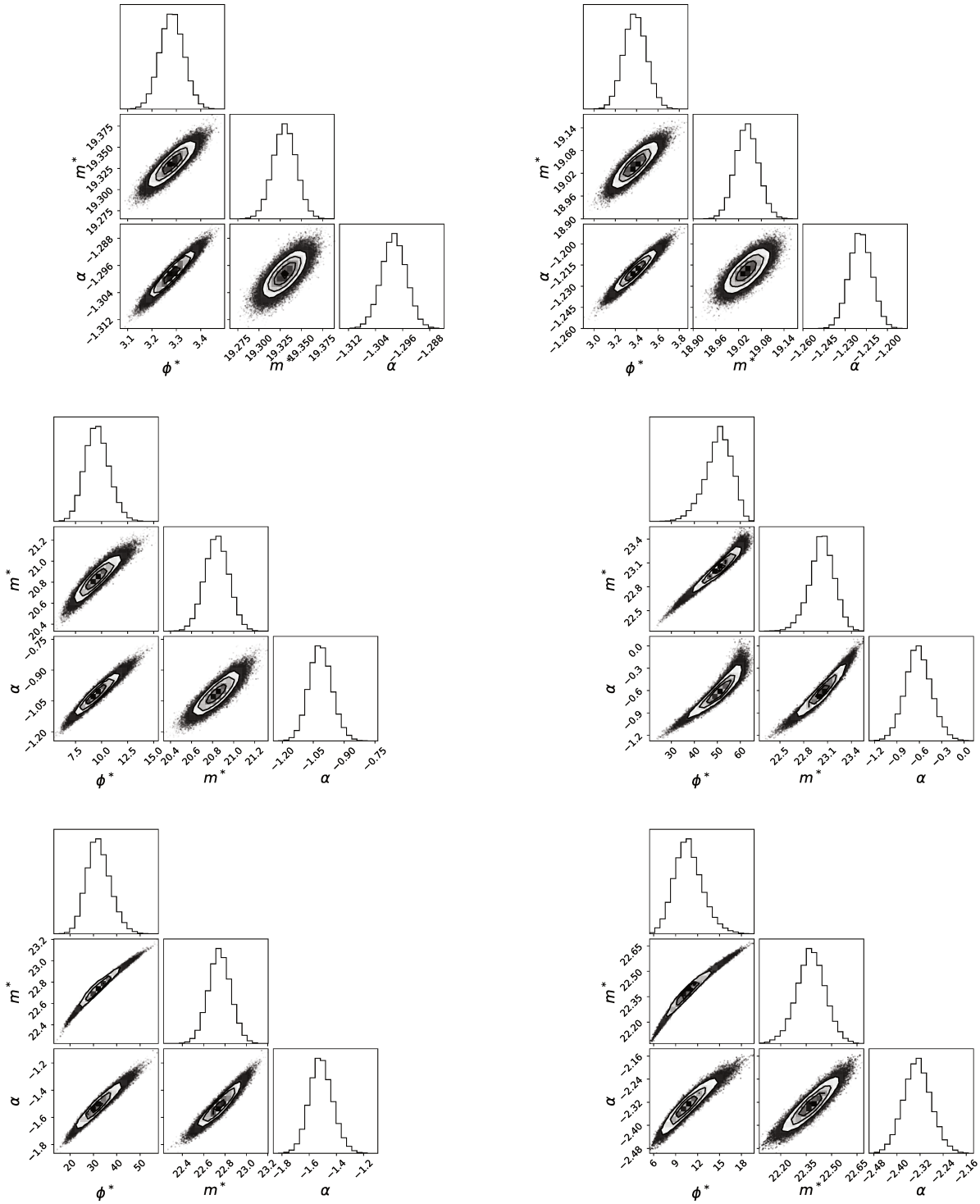


Figure 5.13 – Corner plots of the LFs (black points) fitted to a Schechter model. For the upper row the bins are: in redshift is $0.3 < z < 0.6$ and in mass $10^{13.13}\mathrm{M}_\odot < \mathrm{M} < 10^{13.27}\mathrm{M}_\odot$ (left) and in redshift is $0.3 < z < 0.6$ and in mass $10^{14.74}\mathrm{M}_\odot < \mathrm{M} < 10^{14.88}\mathrm{M}_\odot$ (right). The middle row represents the bins: in redshift is $0.6 < z < 1$ and in mass $10^{15.15}\mathrm{M}_\odot < \mathrm{M} < 10^{15.28}\mathrm{M}_\odot$ (left) and in redshift is $1.3 < z < 1.6$ and in mass $10^{14.6}\mathrm{M}_\odot < \mathrm{M} < 10^{14.74}\mathrm{M}_\odot$ (right). The bottom row represents the bins: in redshift is $1.6 < z < 2$ and in mass $10^{14.2}\mathrm{M}_\odot < \mathrm{M} < 10^{14.34}\mathrm{M}_\odot$ (left) and in redshift is $2 < z < 2.3$ and in mass $10^{13.53}\mathrm{M}_\odot < \mathrm{M} < 10^{13.67}\mathrm{M}_\odot$ (right).

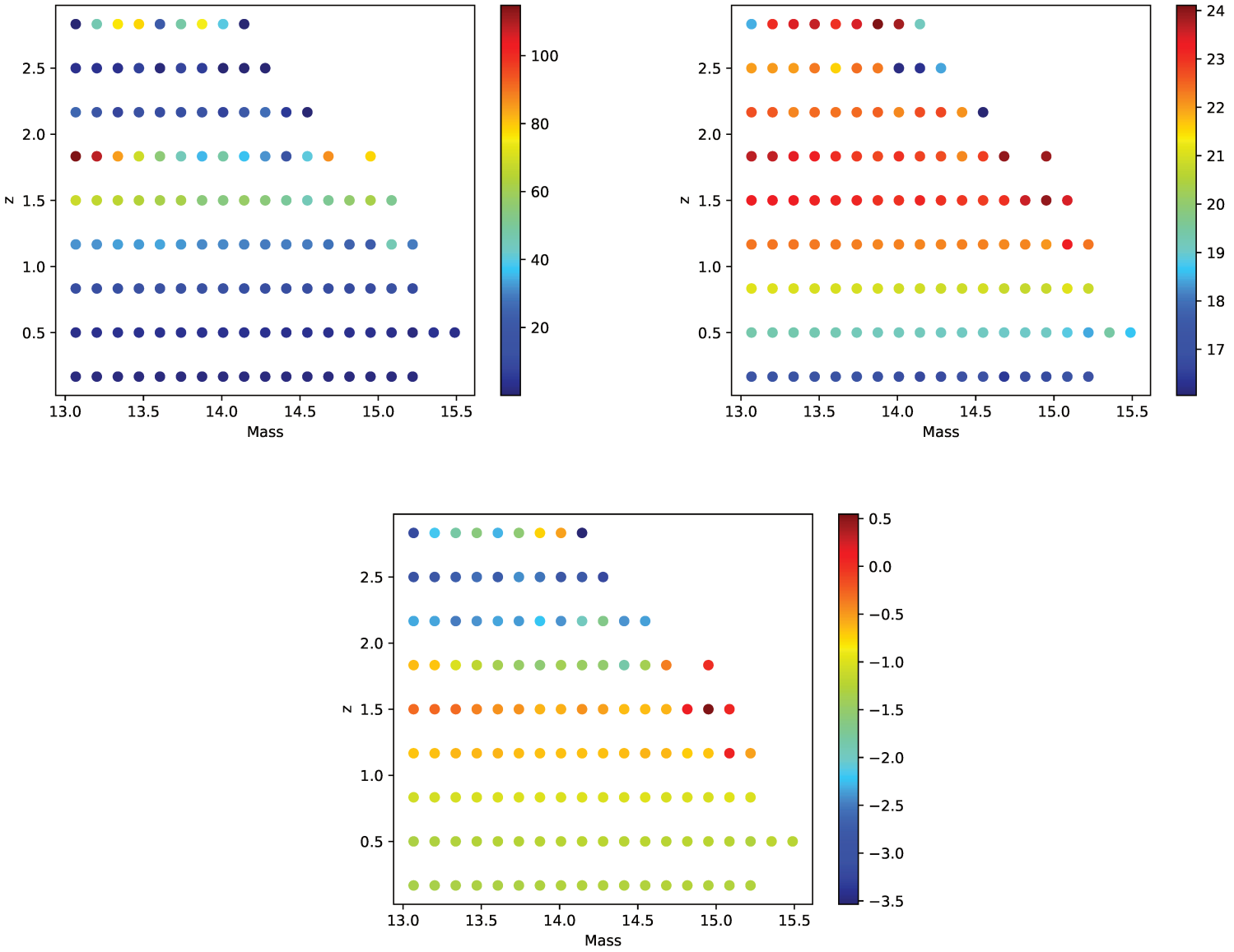


Figure 5.14 – Color coded best-fit parameters, $\phi^*$, m$^*$ in the upper row, respectively, and $\alpha$ in the bottom row, as a function of mass and redshift.

Chapter 6

# Constructing a Synthetic Cluster Catalogue



In this chapter we construct a synthetic cluster catalogue based on the *Euclid* Mock catalogue properties. First, we present the methodology followed to construct the synthetic catalogue from analytical models, in particular: the galaxy density distribution, and the luminosity function. Second, we recover the cluster properties once the catalogue is constructed. Finally, we compare some of these properties with the *Euclid* mock catalogue.

## 6.1 Methodology

The cluster injection method requires a catalogue with simulated synthetic clusters that is constructed from the main properties of the reconstructed cluster catalogue. In the case of *Euclid*, there is no real data, so we work with galaxy mock catalogues. Using the galaxy mock catalogue and the results obtained in Chapter 5, we simulate galaxy clusters with a NFW [101] radial galaxy density distribution and a Schechter model [93] for the luminosity function (LF).

Figure 6.1 shows a diagram of the process for constructing the synthetic cluster catalogue. The computational cost to construct a catalogue of 300 deg$^2$ is high (size of the *Euclid* Mock

catalogue), thus we decided to construct a smaller portion of about 36 deg$^2$. We choose a region in sky coordinates within the limits of the *Euclid* Mock catalogue: 14.3° < RA < 20.5° and -4.1° < DEC < 2°. We use a threshold in mass and redshift: M > $10^{14}$ $M_\odot$ and z < 2.

We recover the observational properties studied in Chapter 5 for each bin in mass and redshift. First, we compute the number of clusters in each bin. Because the number of clusters is related to a catalogue of 300 deg$^2$, we extrapolate to 36 deg$^2$.

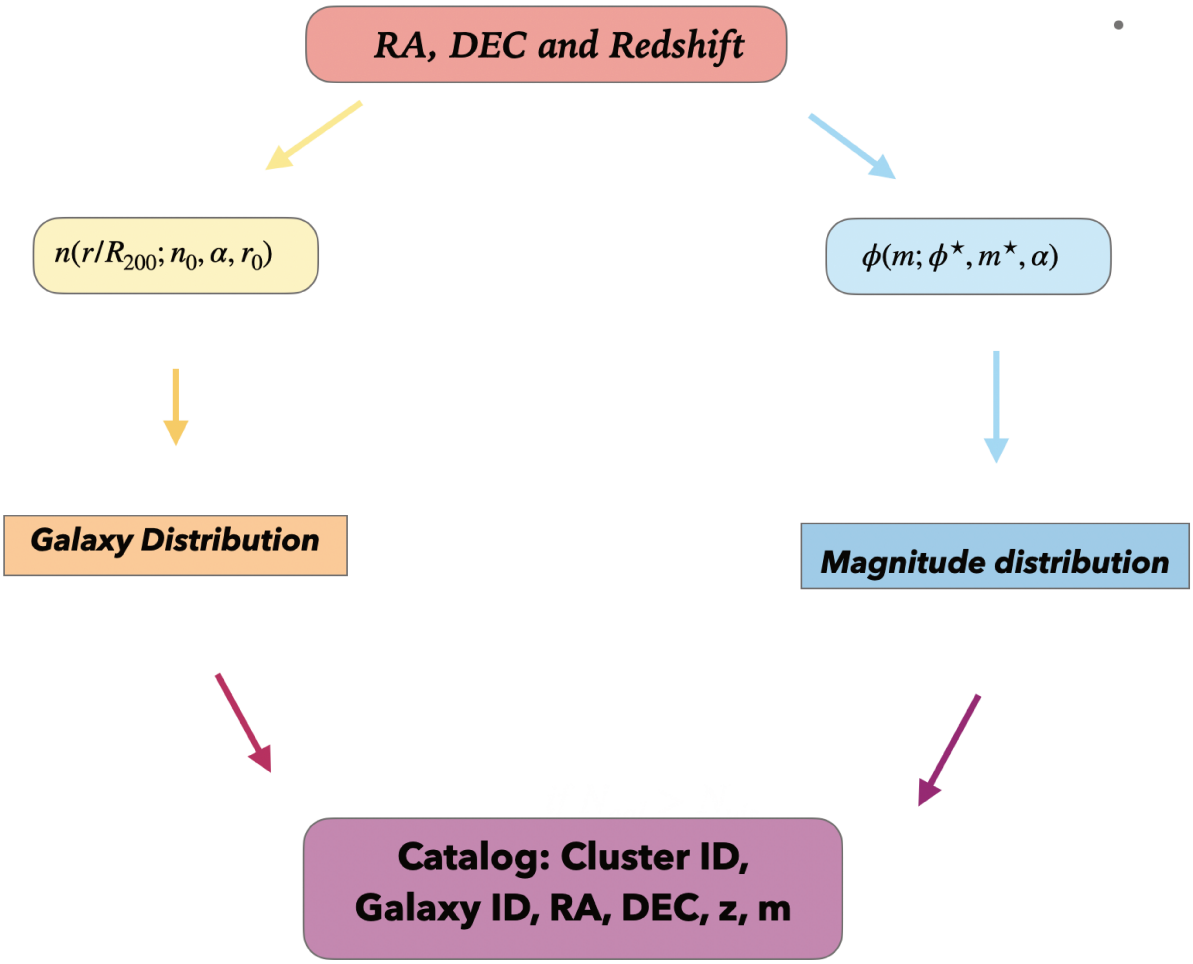


Figure 6.1 – Scheme of the methodology for constructing a synthetic cluster catalogue.

For each cluster in the bin we calculate the number of cluster member in two ways:

1. We compute the 3D NFW profile using the best-fit parameters obtained in Chapter 5 for the corresponding bin in mass and redshift. We project in 2D this profile following the procedure explained in Chapter 5. The number of cluster members as a function of

the distance are computed in a differential way, thus the projected radius is separated in bins. We then integrate the 2D projection over the projected radius to compute the total number of member galaxies.

2. We compute the LF with a Schechter model for the corresponding bin in mass and redshift using the best-fit parameters obtained in Chapter 5. The number of cluster member galaxies is calculated by integrating the LF.

The number of cluster member galaxies can differ between the first and the second method, because the results depend on the quality of the fits discussed in Chapter 5. Therefore, we decide to choose the minimum value of cluster members between the two methods.

For each bin in mass and redshift we have computed the number of clusters for a region of 36 deg $^2$ and then we simulate each cluster as follows. We first assign to each cluster in the bin, a mass and a redshift randomly with a uniform distribution within the bin limits. Then, we compute the number of galaxy members in the cluster and assign to each of them sky positions, redshifts and magnitudes as explained bellow:

- **Cluster Sky Position:** The position of the cluster on the sky is draw randomly following an uniform distribution. Notice that in this way the spatial correlation between clusters is not preserved.
- **Sky Position of Cluster's Member Galaxies:** We generate the 3D galaxy density distribution following a NFW model using the parameters corresponding to the cluster bin in mass and redshift. This distribution is projected to obtain a radial 2D profile as explained in Chapter 5. For each radial bin in the obtained 2D profile, we compute the number of galaxy members as discussed above. Thus, we simulate those galaxies distributing randomly in a spherical shell at that radius. Notice that in our methodology all the clusters in the same bin in mass and redshift will have the same number of galaxy members. In reality this should not be the case as the clusters have different masses and redshift. The final position of each galaxy is obtained by adding the cluster position.
- **Redshift of Cluster's Member Galaxies:** The cluster member galaxies' redshifts are assigned from the cluster redshift adding gaussian uncertainties, i.e., $z_{gal} = z_{cl} \pm \sigma_z$. The standard deviation of the photometric redshifts with respect to the true redshifts in the context of *Euclid* is required to be $\sigma_z/(1+z) < 0.05$. Thus we select $\sigma_z = (1+z_{cl})$ x 0.05. The redshift distribution of the cluster members is a Gaussian distribution with mean $z_{cl}$ and dispersion $\sigma_z$.
- **Magnitudes of Cluster's Member Galaxies:** To assign a magnitude distribution to the galaxies we should use the Schechter model using the best-fit parameters. However, in Chapter 5 we observed that the Schechter model can not reproduce the brightest part of

the luminosity function. This can affect significantly the performance of cluster finders. Therefore we have decided to use the luminosity function of the Euclid Mock Catalogue for sampling the luminosity function of the synthetic catalogue. For each magnitude bin in the obtained LF, we compute the number of galaxy members as discussed above. Thus, we draw randomly the magnitude for each member galaxy from the binned LF. All the clusters in the same bin in mass and redshift will have the same number of galaxy members.

- **Brightest Central Galaxy (BCG):** The brightest galaxy of a cluster is located approximately in its center. Therefore, we assign the central sky position (RA,DEC) in degrees to the brightest galaxy (brightest magnitude draw from the distribution).

To finish the construction of the synthetic cluster catalogue we assign to each cluster member two identification numbers: an individual galaxy ID, and a cluster ID, which is the same for each galaxy belonging to the same cluster. Because we extract the data from the *Euclid* Mock Catalogue we have access to the true cluster mass, which is not an observable and it will not be used when performing a cluster finder in the synthetic catalogue. This is the reason why the mass does not appear in Figure 6.1. Finally we obtain a synthetic cluster catalogue containing the cluster IDs, the galaxy IDs, sky coordinates (RA,DEC) in degrees, redshift, magnitude, and cluster mass.

From the synthetic catalogue we extract a *halo synthetic catalogue* with information only on the clusters. This catalogue has one row per cluster and the information is the one corresponding the BCG. This catalogue will be used in Chapter 7 for matching the clusters to the ones detected by the cluster finder algorithm.

## 6.2 Cluster's in the Synthetic Catalogue

Figure 6.2 shows the sky coordinates in degrees RA, DEC of the member galaxies of a simulated cluster of galaxies of the Synthetic Cluster Catalogue. Each dot represents a galaxy and the colorbar its apparent magnitude in the H-band. The cluster has a mass of M = 10 x $10^{15.1}$ $M_\odot$ and a redshift of z = 0.7. We observe a higher concentration of galaxies in the center of the cluster. In terms of the magnitude distribution, most of the galaxies are faint. These results are in agreement with what is shown in Chapter 5.

We have simulated a total of 2580 synthetic clusters of galaxies in the mass and redshift bins: $10^{14}$ $M_\odot$ < M < $10^{15.45}$ $M_\odot$ and 0 < z < 3. The cluster distribution as a function of the mass and redshift is shown in figure 6.3. Clusters are concentrated at low mass around redshift one. These distribution is in agreement with Figure 5.1 of Chapter 5. However, we observe clusters

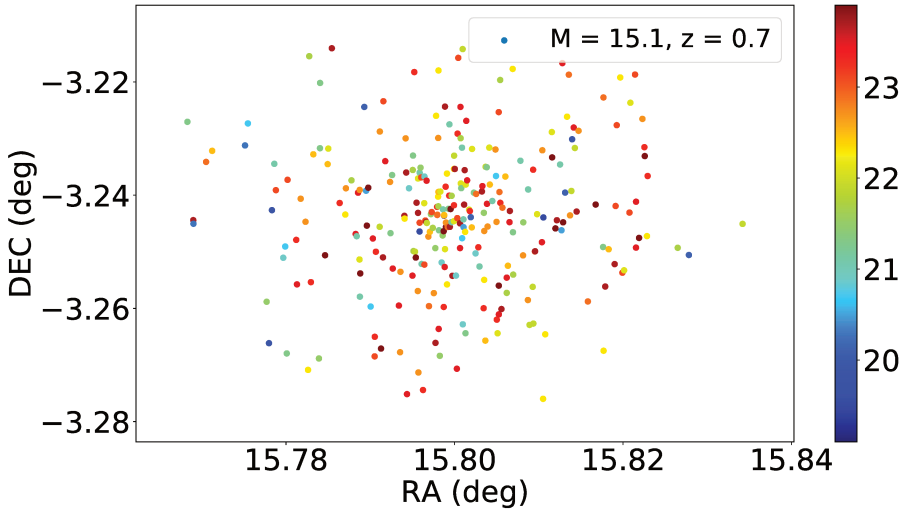


Figure 6.2 – Simulated synthetic cluster of galaxies from the analytical models. Each dot represents a galaxy and the colorbar its apparent magnitude in the H-band. The cluster has a mass of M = 10 x $10^{15.1}$ $M_{\odot}$ and a redshift of z = 0.7.

at more than redshift z = 3, which is the limit of the *Euclid* Mock catalogue. This is due to the simulated uncertainties in the redshift of the cluster member galaxies. Nevertheless, there is only one cluster in this case, and for the results in Chapter 7 we will only consider the results z $<$ 2, as used for cosmological analysis.

The distribution of the cluster member galaxies for the full synthetic catalogue in terms of their sky coordinates (RA,DEC) is shown in Figure 6.4. The colors represent the number density of galaxies. The yellow dots correspond to galaxy overdensities at the cluster positions.

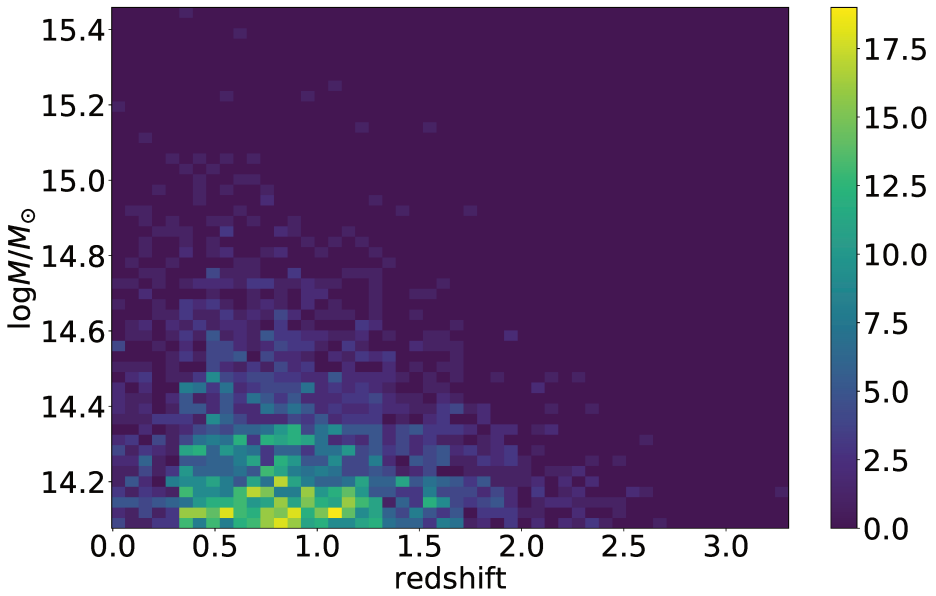


Figure 6.3 – 2D histogram of the distribution of cluster of galaxies in the synthetic cluster catalogue of 36 deg$^2$ as a function of the mass and redshift.

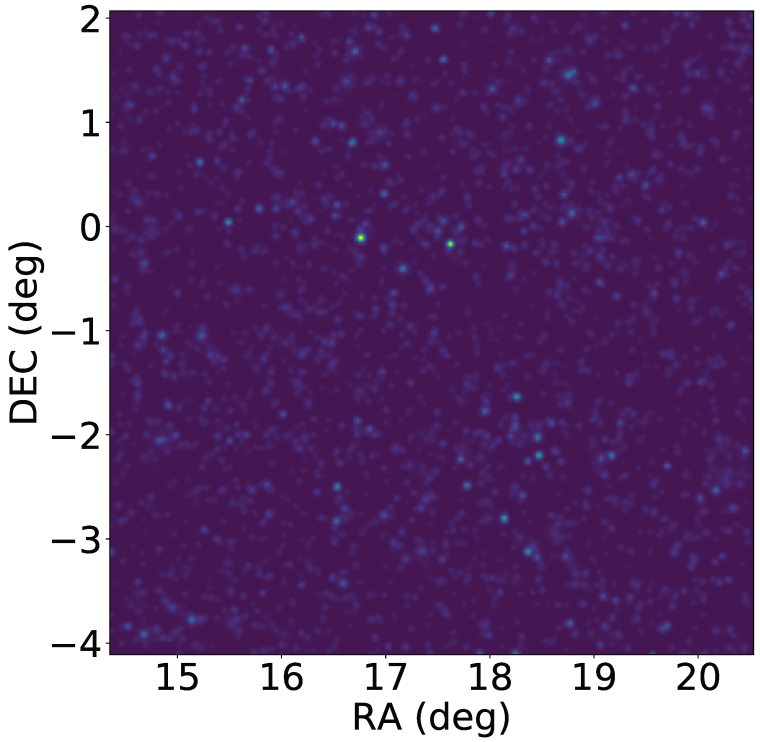


Figure 6.4 – 2D histogram of the number of galaxies for the synthetic cluster catalogue of 36 deg$^2$. The yellow dots correspond to galaxy overdensities at the most massive cluster positions.

## 6.3 Comparing Observational Properties with the *Euclid* Mock Catalogue

To compare the synthetic cluster catalogue to the *Euclid* Mock Catalogue we selected a portion of 36 $deg^2$ with the same limits of the synthetic cluster catalogue. Figure 6.5 shows a 2D histogram in mass and redshift of the number of clusters of the *Euclid* Cluster's Mock Catalogue. We observe that clusters are concentrated at low mass at around redshift one, as we observed for the synthetic cluster catalogue in figure 6.3. We present several examples of clusters of galaxies in figure 6.6 as we did in figure 6.2. The left column corresponds to clusters of the *Euclid* Mock catalogue and the right column shows simulated clusters of the synthetic cluster catalogue. The cluster of galaxies presented in each row present similar mass and redshift for both catalogues:

- **First Row:** Clusters of galaxies of mass $M = 10^{15.1}$ $M_\odot$ at redshift $z = 0.3$ and $z = 0.2$ for the *Euclid* Mock catalogue and synthetic cluster catalogue, respectively.
- **Second Row:** Clusters of galaxies of mass $M = 10^{15.3}$ $M_\odot$ and $M = 10^{15.2}$ $M_\odot$ at redshift $z = 0.9$ and $z = 1$ for the *Euclid* Mock catalogue and synthetic cluster catalogue, respectively.
- **Third Row:** Clusters of galaxies of mass $M = 10^{14.1}$ $M_\odot$ at redshift $z = 1.4$ and $z = 1.3$ for the *Euclid* Mock catalogue and synthetic cluster catalogue, respectively.
- **Fourth Row:** Clusters of galaxies of mass $M = 10^{14.1}$ $M_\odot$ at redshift $z = 0.3$ for the *Euclid* Mock catalogue and synthetic cluster catalogue, respectively.

We observe a general agreement in terms of the galaxy magnitude distribution. However, for the second row and for fourth row, the magnitude distribution is different for both catalogues. In the second row, there are brighter galaxies for the *Euclid* case. In the fourth row, the cluster presents fainter galaxies. Nevertheless, following the color code, it looks like the number of these extreme cases of bright galaxies is negligible. This is in agreement with the results in Figure 6.8, which will be presented later. For the number of galaxies, massive clusters have more cluster members, as we expected. The synthetic galaxy clusters tend to be more relaxed (spherical) than the Euclid ones, because we distributed the galaxies circularly. At low redshift clusters present more galaxies than at high redshift, as one would expect. In general, galaxy clusters of the synthetic catalogue are in agreement in number of galaxies with those of the *Euclid* Mock catalogue. These results are in agreement with the richness distribution presented in figure 6.7. Each dot represents a cluster, in red in the case of the synthetic cluster catalogue and in blue for the *Euclid* Mock catalogue. The top panel corresponds to the richness (number of cluster members) as a function of the cluster mass, while the bottom panel corresponds to the richness as a function of the cluster redshift. The top panel is consistent with figure 5.2 from Chapter 5. We observe horizontal lines in the figure corresponding to cluster of the same bin in mass and

redshift. This is due to the fact that in the synthetic cluster catalogue we considered the same number of galaxy members for clusters in the same bin in mass and redshift. Furthermore, the number of cluster members decreases with redshift and increases with mass, as previously observed in Figure 6.6. Overall, the richness distribution in mass and redshift for both catalogues is consistent.

The next key property we want to recover is the luminosity function, because cluster finders are, in general, sensitive to the brightness of galaxies. Figure 6.8 shows the distribution of the H-band apparent magnitudes, $m_H$, for each catalogue: in blue the *Euclid* Mock catalogue, and in orange the synthetic cluster catalogue. Because we have chosen the LF of the *Euclid* Mock catalogue for sampling the LF of the synthetic clusters, both distributions are consistent. However, we see a bias in the magnitude towards fainter values and also a decrease in the number of galaxies in the brightest region. This may be due to the way we randomly draw the magnitude of the galaxies from the discretized LF. We decided not to take into account clusters with z > 2 in the next chapter because the bin used to produce the synthetic catalogue is large, and we have little statistics.

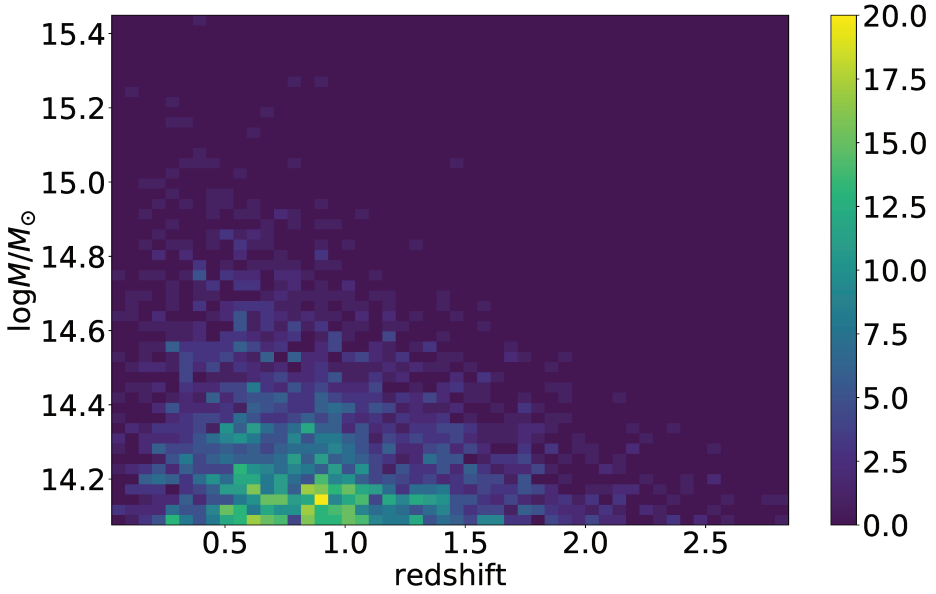


Figure 6.5 – 2D histogram of the distribution of cluster of galaxies in the Euclid Mock Catalogue of 36 deg$^2$ region as a function of the mass and redshift. This distribution can be compared to that of the synthetic catalogue in Figure 6.5.

## 6.4 Conclusions

In this chapter we presented the construction of a synthetic cluster catalogue based on the *Euclid* Cluster's Mock catalogue properties as discussed in Chapter 5.

Overall we see that the cluster properties of the synthetic cluster catalogue are consistent with the *Euclid* Mock catalogue. However, there are some aspects that can be improved when simulating the synthetic clusters. First, the number of cluster's galaxies in a bin in mass and redshift is taken as constant. We could have considered a measurement of the dispersion across the bin accounting for a Poisonnian uncertainty. Second, the bins in mass and redshift should probably be narrower to ensure similar cluster properties for each cluster in the bin. Third, we draw magnitudes and galaxy positions from discretized functions, and this leads to some kind of bias as we observed for the luminosity function. Fourth, we consider a spherical shape for clusters, while we observe that clusters in the *Euclid* Mock catalogue present different types of morphologies. We could try to consider a triaxial modelling of clusters that takes into account ellipticity.

Finally, we must notice that for generating synthetic cluster we have used the properties of the *Euclid* Mock catalogue for which we have the "true" mass, redshift, LF and galaxy distribution. In a more realistic case we will only have access to the properties of detected for which will only have estimates of the mass and redshift from the cluster finders and a limited number of galaxy members to derive the LF and the galaxy density distribution. How to go from detected cluster observational properties to "realistic" properties of the synthetic clusters need to be investigated further.

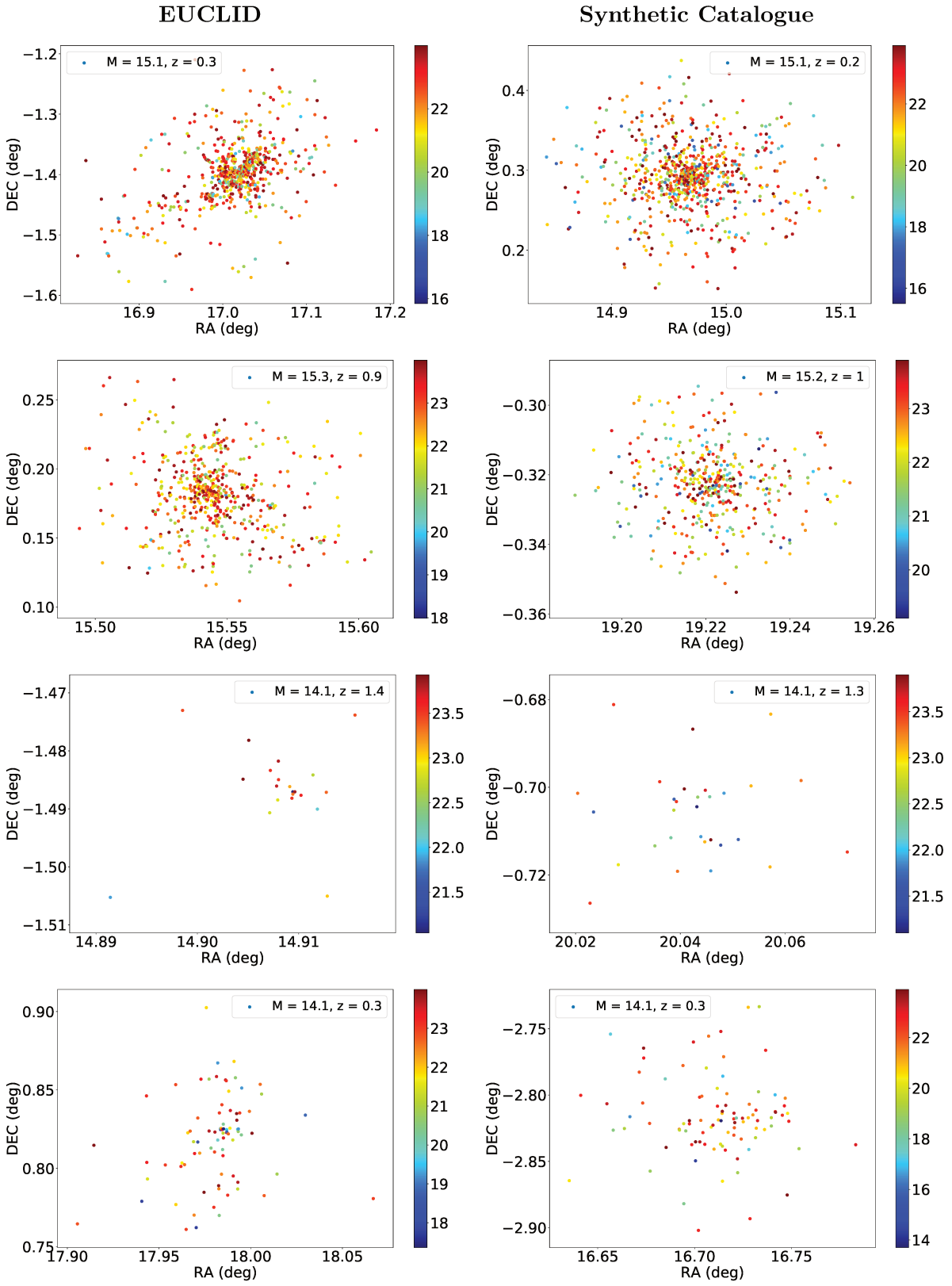


Figure 6.6 – Distribution in RA,DEC of cluster member galaxies from the *Euclid* Mock catalogue and the synthetic cluster catalogue, in the left and right columns, respectively. Each dot represent a galaxy and the colorbar their apparent magnitude value in the H-band. Each row represents a bin in mass and redshift, similar for both catalogues. See description of the bins in Section 6.3.

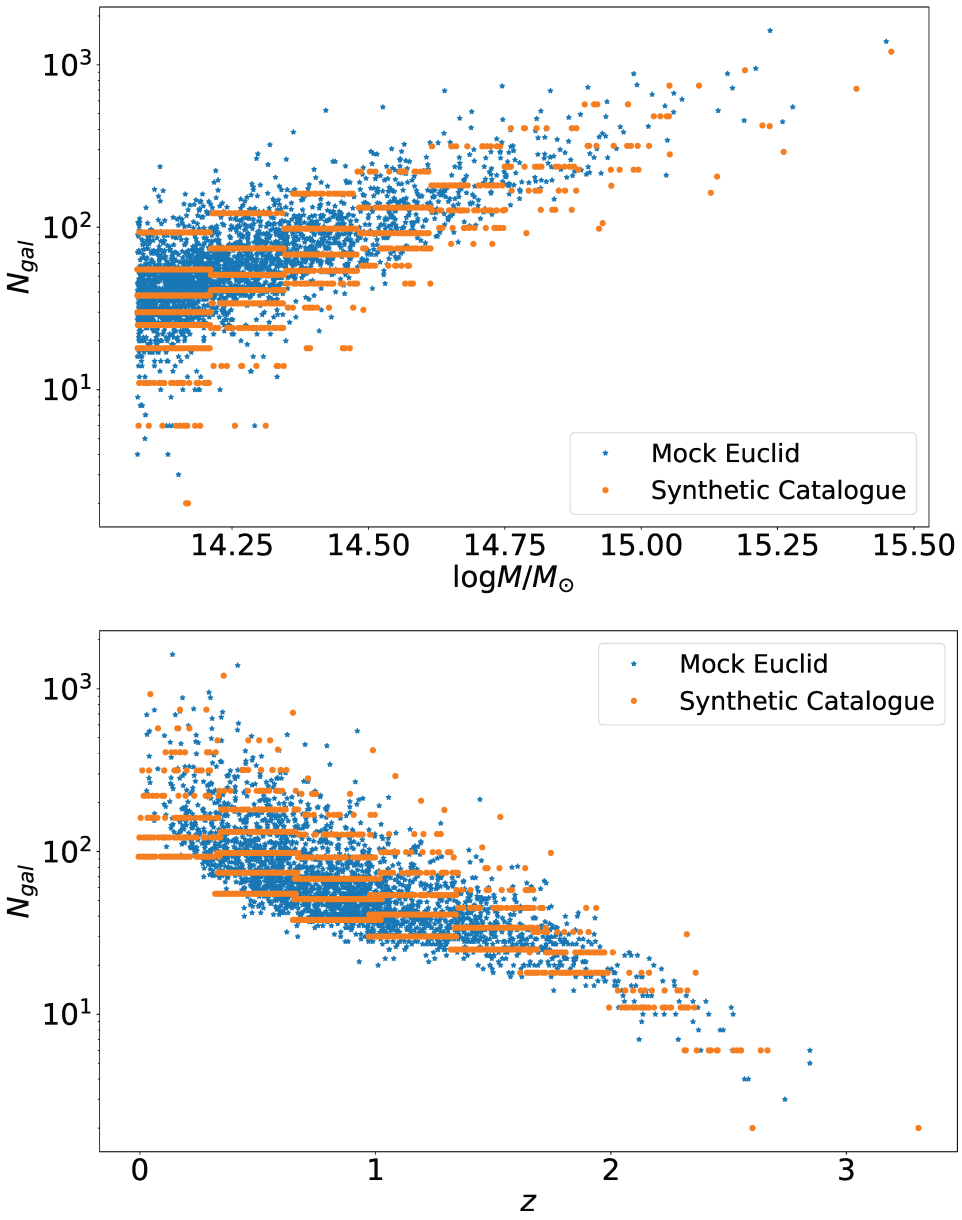


Figure 6.7 – Distribution of the number of galaxies (richness) as a function of the cluster mass (top panel) and cluster redshift (bottom panel). The blue and orange points represent galaxy clusters for the *Euclid* mock catalogue and the synthetic cluster catalogue, respectively.

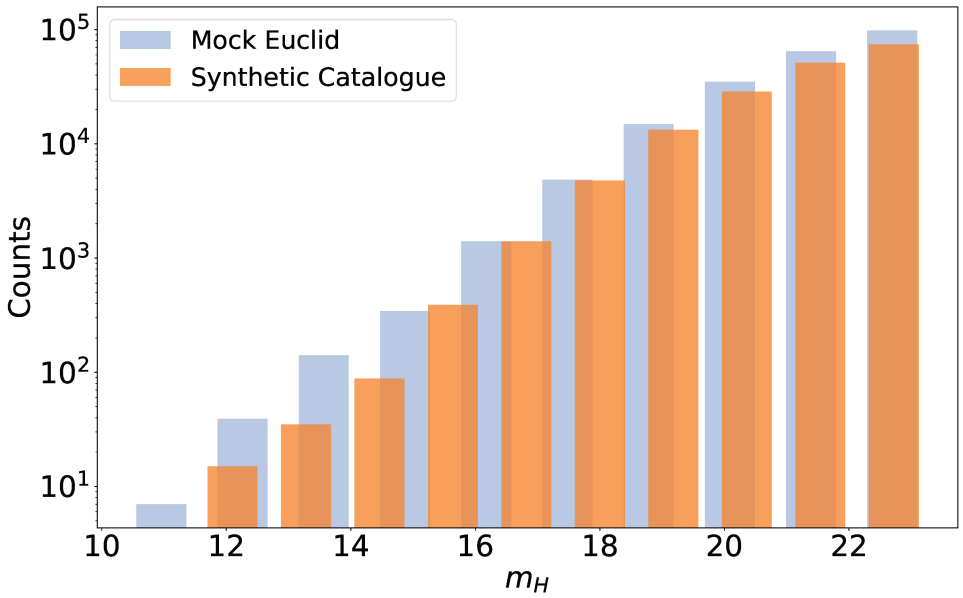


Figure 6.8 – Distribution of the H-band apparent magnitudes for the cluster galaxy members, for the *Euclid* mock catalogue (blue) and the synthetic cluster catalogue (orange).

Chapter 7

# Cluster Injection for Computing the Selection Function



The cluster abundance is related to the parameters of the $\Lambda CDM$ cosmological model. But to be able to recover the total number of clusters in a solid angle region, $\Omega$, it is necessary to account for the probability of finding a cluster at a certain mass and redshift in this solid angle (see Chapter 2). This probability is known as the Selection Function and it is an intrinsic characteristic of the studied cluster catalogue and would depend on the cluster finder selected, the observational and quality cuts, and the survey. In this chapter we study the performance of the PZWAV cluster finder on several catalogues constructed using the synthetic cluster catalogue presented in Chapter 6. For reference we will also use the 36 deg$^2$ *Euclid* Mock catalogue. First, we present the catalogues we will use during the chapter. Second, we introduce the PZWAV cluster finder, which is officially accepted by the *Euclid* Consortium. Third, we apply the PZWAV algorithm individually on each catalogue, and we match the input cluster (catalogues) with the output ones (PZWAV) by a geometrical matching method. From this, we compute the completeness and purity and we study the dependency with the geometrical parameters of the matching procedure. Finally, we compare the completeness and purity for the fourth catalogues and the richness estimates from the PZWAV algorithm.

## 7.1 Galaxy Catalogues

To compute the selection function of the *Euclid* Mock catalogue we have decided to use four catalogues: the 36 deg$^2$ *Euclid* Mock catalogue, the synthetic cluster catalogue discussed in chapter 6 with field galaxies, and two injection catalogues which combine properties of the previous two.

### 7.1.1 Synthetic Galaxy Catalogue

To have a full independent galaxy catalogue, we add field galaxies to the synthetic cluster catalogue constructed in Chapter 6. For computing the number of galaxies to be added, we have counted the number of galaxies in the 300 deg$^2$ *Euclid* mock catalogue, and we extrapolated to the 36 deg$^2$ region. We need to distribute the field galaxies in sky coordinates, redshift and magnitude. The field galaxies have been spatially distributed uniformly within the limits of the 36 deg$^2$ catalogue. For the redshift distribution we use the below probability distribution function [187]:

$$P(z) = z^{\alpha} \exp\left[-\left(\frac{z}{z_0}\right)^{\beta}\right], \tag{7.1}$$

with $\alpha = 1.24$ and $\beta = 1.01$ and $z_0 = 0.51$. As for the synthetic cluster catalogue the magnitude for each galaxy has been draw from the LF of the *Euclid* Mock catalogue. However, we have sampled the luminosity function coming from the Schechter fit done in Chapter 5, instead of directly the LF of the *Euclid* Mock Catalogue, as done in Chapter 6. From now on this catalogue will be called *synthetic catalogue*.

### 7.1.2 Injection Galaxy Catalogues

To compute the selection function using the cluster injection method (see Chapter 2), we have to construct an injected galaxy catalogue. To do this, we propose two methods in the following.

#### Euclid Mock Catalogue + Synthetic Cluster Catalogue

In the first method, we add the galaxies of synthetic clusters directly in the *Euclid* Mock catalogue, without any other requirement. This creates a catalogue that keeps the environment spatial correlations, but add an excess of clusters with respect to the *Euclid* Mock catalogue. The synthetic cluster can overlap with the existing ones. This catalogue allows us to check the performance of cluster finders when there are clusters with different nature (simulated in a different way), and the effects of cluster's overlapping. From now on this catalogue will be called *Injection Catalogue*.

#### *Euclid* Mock Galaxies + Synthetic Cluster Catalogue

In the second method, we randomized the galaxy positions and redshifts in the *Euclid* Mock catalogue. This removes the mock galaxy clusters and the spatial correlations for both the environment and the clusters. We add galaxies of the synthetic clusters directly in this catalogue. With this method we preserve the number of galaxies and remove the cluster. Notice that in

practice, the galaxies from the clusters in the *Euclid* Mock catalogue will add to the field galaxies. It would be better to try to "remove" the clusters and so preserve both the distribution of field galaxies and their spatial distribution. This catalogue allows to check the performance of cluster finders with respect to the environment. From now on this catalogue will be called *Synthetic Catalogue with Euclid 's randomized galaxies.*

## 7.2 *Euclid* Cluster Finders: PZWAV

In the context of *Euclid* the test of several cluster finders has been done in four Cluster Finder Challenges (CFC) between 2013 and 2017. In Adam et al. [134] they concluded that two cluster finders give best results and can be implemented in the *Euclid* pipeline: the Adaptative Matched Identifier of Clustered Objects (AMICO) code [133] (see Chapter 1) and the PZWAV code [188] (the one we use in this thesis).

PZWAV searches for overdensities in galaxy density maps using a wavelet approach [127] that has been optimized for *Euclid*-like data. The steps to follow are:

- Prepare a galaxy catalogue with photometric redshifts, sky coordinates, magnitudes and the probability distribution associated with each photometric redshift, P(z), that gives us the probability of the galaxy to be found at a redshift $z$.

- The algorithm constructs galaxy density maps placing the galaxies in redshift slices weighted by P(z), that cover the *Euclid* redshift range.

- The galaxy density maps are then convolved with a difference-of-Gaussians smoothing kernel of a fixed physical size, which is comparable with the cluster cores' physical size.

- Then, clusters are detected as peaks in the wavelet smoothed galaxy density maps for each redshift slice. The detections are merged across the redshift slices.

- The density peaks is the direct observable of the search and it can be taken as a proxy of the richness. PZWAV also computes the SNR of the detection. The cluster redshift is estimated from the median and the standard deviation of the distribution of photometric redshifts for all the galaxies that lies at $30''$ from the cluster core and inside $\Delta z = 0.12$ of the redshift slice where the cluster is.

- PZWAV produces a cluster catalogue with estimates of the photometric redshift, sky

coordinates, richness and SNR for each of the detected clusters.

In the following we decided not to use the SNR but the percentage in detections after ordering them by decreasing SNR. Because we have different types of catalogues, a cut in SNR does not include the same amount of clusters for both catalogues. A cut in the percentage of detections will ensure a more homogeneous selection. As proposed in Adam et al. [134] the cuts in percentages of detections that we will use are: 1%, 5%, 10%, 20%, 40%, 80% and 99%, where 1% considers only the highest SNR detections and 99%, considers almost every detection.

## 7.3 Completeness and Purity of the Simulated Catalogues

The performance of a detection algorithm depends on the quality of the cluster catalogue that it produces. We need to know how many of the existing clusters it detects, called completeness, and how many of them are real, called purity.

Applying the algorithm to a known sample of clusters we define the completeness as the number of clusters that can be associated with a real cluster by the cluster finder, $N_{det}^{true}$, with respect to the ones in the Mock catalogue, i.e.,:

$$C = \frac{N_{det}^{true}}{N_{mock}}. \tag{7.2}$$

In the case of simulated catalogues, the association between the detected cluster and the mock cluster can be one-way or two-way, meaning that, if we compare the number of cluster that are in the finder's cluster catalogue with the number of clusters in the cluster mock catalogue we have a one-way association. If we check how many clusters in the cluster mock catalogue are found in the finder's one, we have another one-way direction. If we compute the common matches for both one-way directions we have what we will call two-way or bijection matches.

The purity, defined as the number of detected "true" clusters as a function of the overall detection number:

$$P = \frac{N_{det}^{true}}{N_{finder}}, \tag{7.3}$$

where $N_{finder}$ is the total number of detections of the cluster finder. Again we can perform the one-way or two-ways purity matches.

In each of the one-way direction matches we can find several cases:

- If a cluster is found more than once in the finder's catalogue, i.e., one single cluster in the cluster mock catalogue is matched with several detections, we say that we have fragmentation.

- If a cluster is found more than once in the cluster mock catalogue, i.e., one single detected cluster is matched with several "true" clusters, we say that we have overmerging.

If the completeness and purity of the bijections is close to the completeness and purity of the one-way matches we can conclude that fragmentations and overmergings are not significant.

### 7.3.1 Geometrical Matching Procedure

For performing the one-way or two-way matching from the input catalogue (the cluster mock catalogue) and the PZWAV detected cluster catalogue, we need to define a procedure to determine when a cluster is matched or not. Following Adam et al. [134], we will use what is called "Geometrical Matching" (that can be applied to any cluster finder):

- A volume around the mock cluster is defined with a depth $D = k\sigma_0(1+z)$. The mean photometric uncertainty for *Euclid* is $\sigma_{zphot} = 0.05(1+z)$, thus $\sigma_0 = 0.05$. To account for the photometric uncertainties in the matching, in Adam et al. [134] they consider $k = 4$. We will study the behaviour of the completeness and purity varying this parameter using $\Delta z = k\sigma_0$. To ensure that every cluster galaxy member is included, and the cluster is well defined, the volume is restricted to the extent of the galaxies belonging to the cluster: $\mathrm{RA}_{min}$, $\mathrm{RA}_{max}$, $\mathrm{DEC}_{min}$ and $\mathrm{DEC}_{max}$. Also, as an additional criteria to the latter, the volume is limited to a distance of $\theta_{200}$ computed, using the catalogue cosmology, from $R_{200}$ of the cluster. The clusters detected inside this volume will be considered matches. However, we will use a fixed value of $\mathrm{R}_{200}$ to compute $\theta_{200}$, as discussed below.

- In case of finding more than one match, the closest detected cluster in projected sky coordinates is selected as the main match, but all the possibilities are recorded to study fragmentation.

- We repeat the last two processes but using the PZWAV detected clusters as a reference. The only difference is that we can not compute $\theta_{200}$ for the PZWAV detected clusters since in general we have no information of the radius $R_{200}$. Therefore, a mock cluster is matched to a PZWAV one if the distance between the two is inside a general fixed $\theta_{MP}$ computed from a given fixed $R_{MP}$ in Mpc, using the catalogue cosmology. Again the excess of matches are recorded as a way of counting for the overmergings.

This method allows us to have a bijection matching since there is the possibility of comparing the matches from both ways associations.

### 7.3.2 Results on the considered catalogues

#### *Euclid* Mock Catalogue

Although Adam et al. [134] already presented results on the PZWAV performance for the 300 deg$^2$ *Euclid* Mock catalogue, we have decided to run the cluster finder on the *Euclid* 36 deg$^2$ catalogue to have a more direct comparison with the results on the synthetic catalogue.

In this section we present the completeness and purity for the PZWAV cluster finder in the 36 deg$^2$ *Euclid* Mock Catalogue. We limit to clusters with M $> 10^{14}$ M$_\odot$ and z $< 2$. From now we will show purity versus completeness as computed for 1%, 5%, 10%, 20%, 40%, 80% and 99% of the detections (ordered in SNR). In addition, we will check the dependency of the purity and completeness with the parameters of the geometrical matching procedure: $\Delta z$ and $R_{MP}$.

Figure 7.1 shows the purity versus completeness for several cuts in detection's percentages, represented as dots, in increasing order from left to right, i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. The top panel shows the values of purity and completeness depending on the geometrical parameter $\Delta z$ for four values: 0.05, 0.1, 0.15 and 0.2 in blue, orange, green and red, respectively, for a fixed value of R$_{MP}$ of 2 Mpc. The bottom panel corresponds to values of purity and completeness depending on the geometrical parameter R$_{MP}$ for two values: 1 Mpc and 2 Mpc, in blue and red, respectively, for a fixed value $\Delta z$ of 0.2.

We observe that both parameters do not affect significantly the matching. Nevertheless, a deeper volume in redshift and $\theta$ increase both purity and completeness. We reached completeness of 70% for a purity of 70% for 5% of the detections. In Adam et al. [134] for the same percentage of detections they reached a completeness of 60% and a purity of 90%. Furthermore, they presented greater values than us for completeness and purity for several cuts in detection's percentage. This can be due to the fact that the parameters we have chosen to run PZWAV differs from the ones used in Adam et al. [134]. Another possible explanation could be that the catalogue we are using is slightly different that the one used in Adam et al. [134] as explained in Chapter 5, or to differences between the 36 deg$^2$ region and the 300 deg$^2$ region.

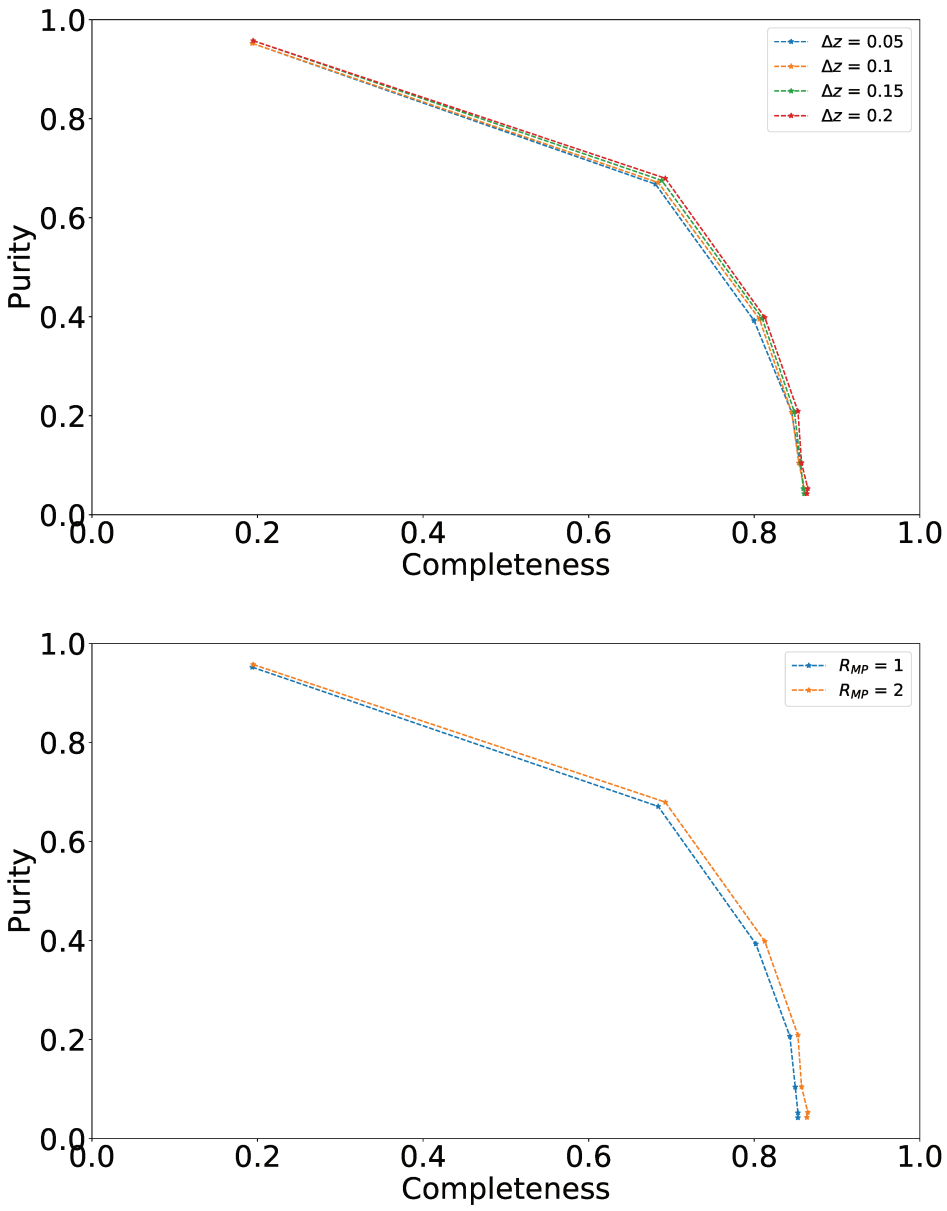


Figure 7.1 – Purity versus completeness for the *Euclid* Mock catalogue, for several cuts in detection percentage, represented as dots, in increasing order from left to right, i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. The top panel shows the values of purity and completeness depending on the geometrical parameter $\Delta z$ for four values: 0.05, 0.1, 0.15 and 0.2 in blue, orange, green and red, respectively, for $R_{MP} = 2$ Mpc. The bottom panel corresponds to values of purity and completeness depending on the geometrical parameter $R_{MP}$ for two values: 1 Mpc and 2 Mpc, in blue and red, respectively, for $\Delta z = 0.2$.

### Synthetic Catalogue

In this section we repeat the previous analysis in the synthetic catalogue. We restrict the analysis to clusters with M $> 10^{14}$ $M_\odot$ and z $< 2$.

Figure 7.2 shows the purity versus completeness for the several cuts in detection percentages. See caption of Figure 7.2 for a detailed description. We observe that both parameters affect the matching. In this case, a deeper volume in terms of redshift does not ensure a better matching. The $\Delta z$ value for which we obtain the best results for purity and completeness is $\Delta z = 0.1$. As we will see in the following results, this can be due to how the clusters are defined and to the differences in the environment. We use as a reference $\Delta z = 0.2$ as did in Adam et al. [134]. We reached completeness of 85% for a purity of 85% for 10% of the detections. This result is more consistent with Adam et al. [134]. In terms of $R_{MP}$, the completeness and purity are greater for a value of 2 Mpc.

The fact that the matching depends on $\Delta z$ and $R_{MP}$ for the synthetic catalogue but not for the *Euclid* one, can be due to the fact that our clusters are spherical and the redshift dispersion is the instrumental one except for *Euclid*. Our clusters tend to be larger in size, and thus a larger volume is needed to detect them.

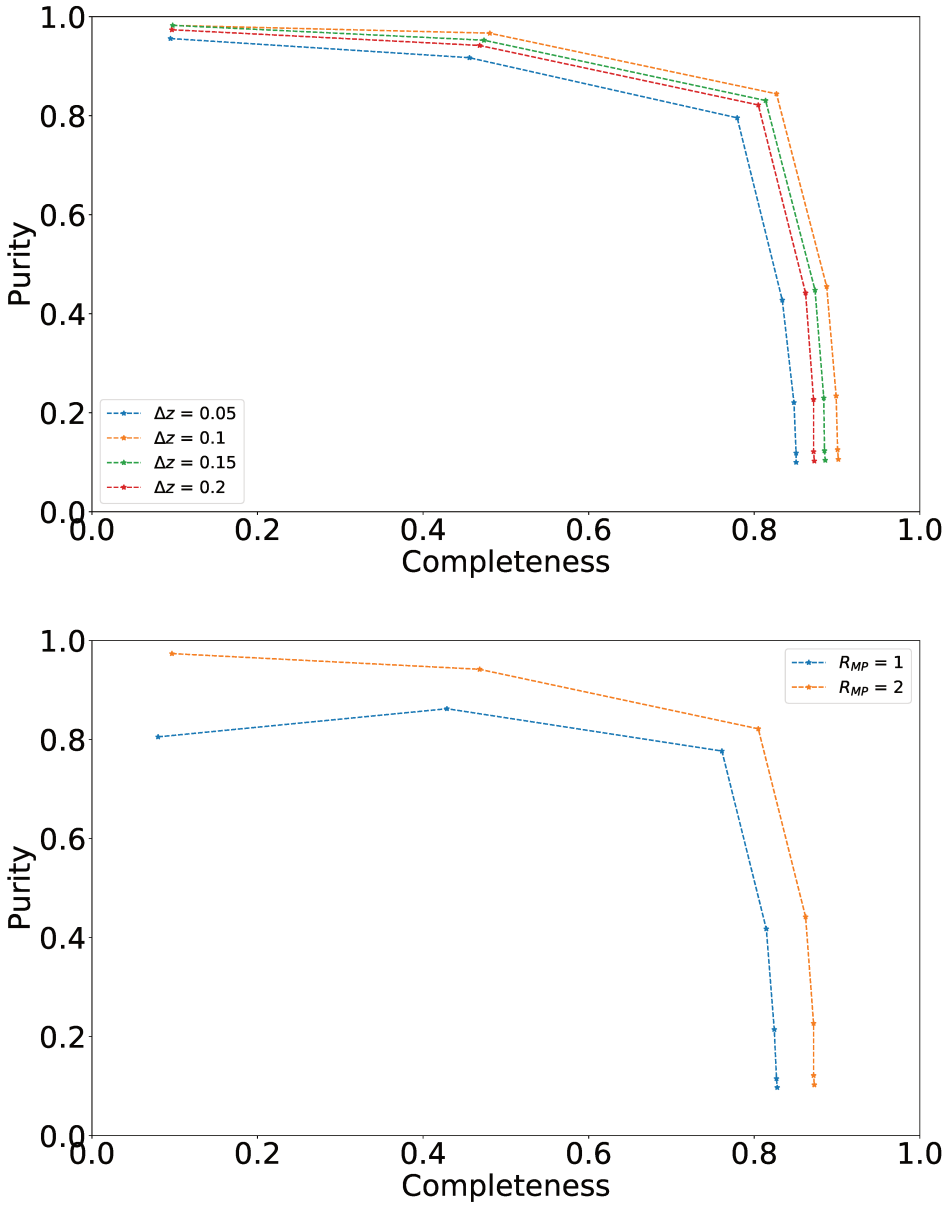


Figure 7.2 – Purity versus completeness for the synthetic catalogue for several cuts in detection percentage, represented as dots, in increasing order from left to right: i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. The top panel shows the values of purity and completeness depending on the geometrical parameter $\Delta z$ for four values: 0.05, 0.1, 0.15 and 0.2 in blue, orange, green and red, respectively, for $R_{MP} = 2$ Mpc. The bottom panel corresponds to values of purity and completeness depending on the geometrical parameter $R_{MP}$ for two values: 1 Mpc and 2 Mpc, in blue and red, respectively, for $\Delta z = 0.2$.

### Injection Catalogue

In this section we repeat the previous analysis in the injection catalogue. We restrict to clusters with M $> 10^{14}$ $M_{\odot}$ and z $< 2$.

Figure 7.3 shows the purity versus completeness for the several cuts in detection percentage. See the caption of Figure 7.3 for more details. In this case, a deeper volume in terms of redshift improve the completeness and purity. We use as a reference $\Delta z = 0.2$. With respect to $R_{MP}$ there is a significant different between 1 Mpc and 2 Mpc. We choose $R_{MP} = 2$ Mpc as reference. We reached completeness of 40% for a purity of 80% for 5% of the detections. Both purity and completeness decrease significantly when increasing percentage of detections.

Comparing with the results of the synthetic catalogue we observe that the choice of parameter $\Delta z$ is affecting less the matching procedure. This can be due to the contribution from the *Euclid* clusters. However, the choice of the parameter $R_{MP}$ changes significantly the results on the completeness and purity. This may be due to the fact that we have an overlapping of the synthetic and *Euclid* Mock clusters.

In addition, clusters are different from both catalogues, thus how clusters are simulated affect cluster finders. PZWAV shows a better performance separately for the *Euclid* catalogue and the synthetic catalogue. Therefore, combining different type of clusters makes difficult to detect clusters.

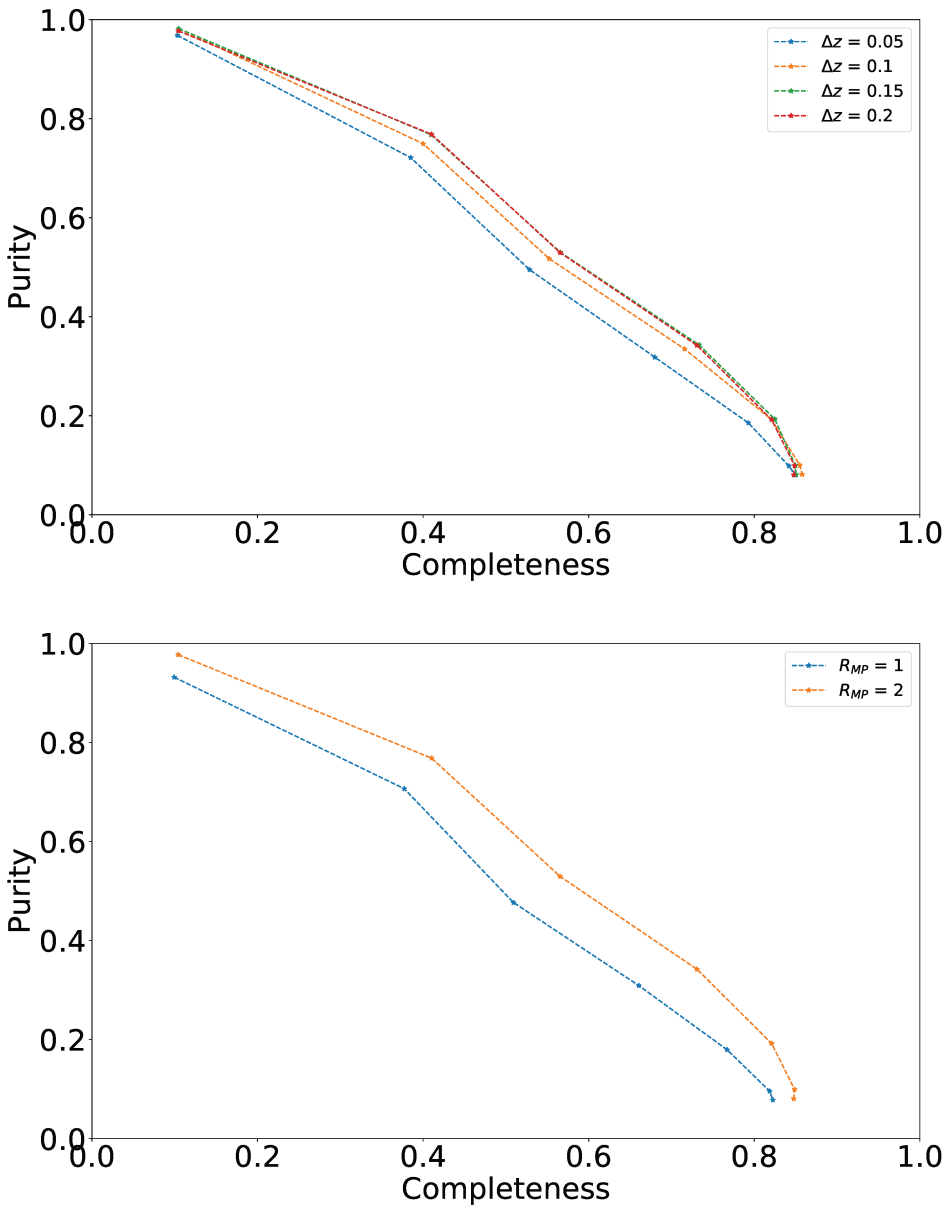


Figure 7.3 – Purity versus completeness for the injection catalogue for several cuts in detection percentage, represented as dots, in increasing order from left to right, i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. The top panel shows the values of purity and completeness depending on the geometrical parameter $\Delta z$ for four values: 0.05, 0.1, 0.15 and 0.2 in blue, orange, green and red, respectively, for $R_{MP} = 2$ Mpc. The bottom panel corresponds to values of purity and completeness depending on the geometrical parameter $R_{MP}$ for two values: 1 Mpc and 2 Mpc, in blue and red, respectively, for $\Delta z = 0.2$.

### Synthetic Catalogue with *Euclid* 's randomized galaxies

In this section we repeat the previous analysis in the catalogue created combining the synthetic clusters with the *Euclid* Mock galaxy catalogue for which the position and redshift of the galaxies were randomized. We restrict clusters with $M > 10^{14}$ $M_{\odot}$ and $z < 2$.

Figure 7.4 shows the purity versus completeness for the several cuts in detection percentage. See caption of figure 7.4 for more details.

We observe that the choice of $\Delta z$ does not affect the matching while $R_{MP}$ does. With respect to the results for the synthetic catalogue, the behaviour with $\Delta z$ tells us that the environment plays a key role in the redshift definition of the cluster from PZWAV. On the other hand, the choice of $R_{MP}$ affects the matching of the same type of clusters when we are changing the environment. Thus, it depends exclusively on the cluster properties. We choose $R_{MP} = 2$ Mpc as reference. We reached completeness of 50% for a purity of 90% for 5% of the detections. The parameter $\Delta z$ affects the performance of the matching for the synthetic catalogue, Figure 7.2, whereas when changing the environment, this dependence disappears.

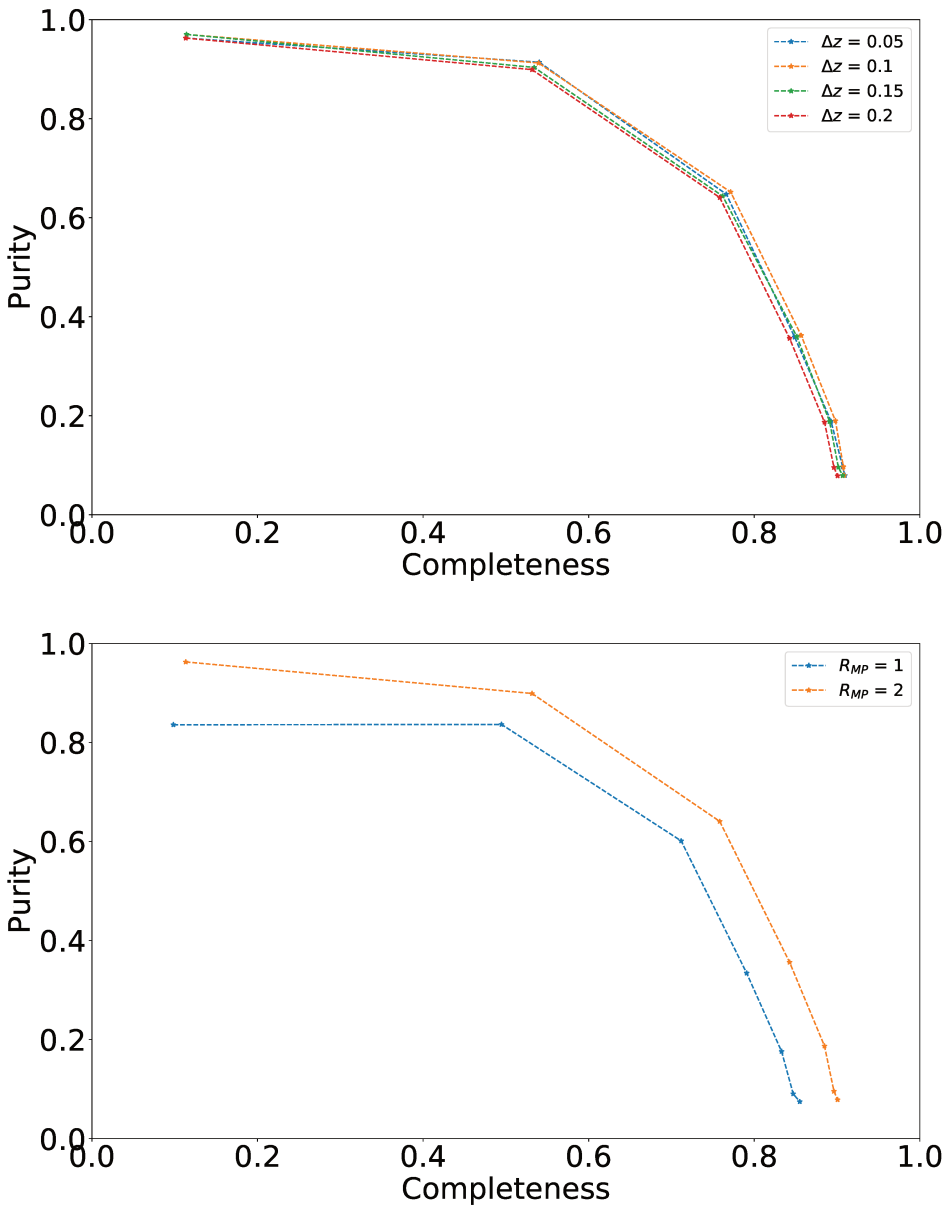


Figure 7.4 – Purity versus completeness for the synthetic catalogue with *Euclid* 's randomized galaxies for several cuts in detection percentage, represented as dots, in increasing order from left to right, i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. The top panel shows the values of purity and completeness depending on the geometrical parameter $\Delta z$ for four values: 0.05, 0.1, 0.15 and 0.2 in blue, orange, green and red, respectively, for $R_{MP} = 2$ Mpc. The bottom panel corresponds to values of purity and completeness depending on the geometrical parameter $R_{MP}$ for two values: 1 Mpc and 2 Mpc, in blue and red, respectively, for $\Delta z = 0.2$.

## 7.4 Comparison of Results

### 7.4.1 Completeness and Purity

Completeness and purity have been computed in the previous section for four cases: the *Euclid* Mock catalogue, the synthetic catalogue, and two injection catalogues: one including both *Euclid*-type clusters and synthetic clusters, and the other including the *Euclid* randomized galaxies and synthetic clusters. We have checked the performance of the matching procedure by studying the dependency of its parameters, $\Delta z$ and $\mathrm{R}_{MP}$, with the completeness and purity. We decided to take as reference $\mathrm{R}_{MP} = 2$ Mpc and $\Delta z = 0.02$. The first is because it improves the matchings, and the second to be consistent with Adam et al. [134].

Figure 7.5 shows the purity versus the completeness for several cuts in detection's percentages, represented as dots, in increasing order from left to right, i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. The blue, orange, red and green curves correspond to the *Euclid* Mock catalogue, the synthetic catalogue, the injection catalogue, and the synthetic catalogue with *Euclid* 's randomized galaxies, respectively. We can compare the results for several cases:

- ***Euclid* Mock catalogue vs Synthetic catalogue:** The performance of PZWAV is significantly different for both catalogues. In the case of synthetic clusters with a simple background where there is no spatial correlation, PZWAV shows better performance in completeness and purity, than in the case of a more complex catalogue. This can be due partly to the difference in shape and size of our clusters, but also to much simpler environmental condition in the synthetic catalogue.
- **Injection Catalogue vs Synthetic Cluster Catalogue:** The worst performance of PZWAV is found for the injection catalogue. The latter conserves the background of the *Euclid* Mock catalogue, but it has an overlapping of clusters. This suggests us that the possible overmerging of clusters with two different natures affects significantly the way of detecting clusters.
- **Synthetic Catalogue with *Euclid*'s randomized galaxies vs Synthetic Catalogue:** In this case we compare two catalogues with the same clusters in the same positions, but changing the environment. We observe how the performance of PZWAV worsens for the overdense environment of the Synthetic Catalogue with *Euclid* 's randomized galaxies. However, the differences in the completeness and purity between these two catalogues are not as significant as in the rest of the cases. For two flat environments the difference in number density of galaxies do not affect the cluster finder.
- ***Euclid* Mock catalogue vs Synthetic Catalogue with *Euclid*'s randomized galaxies:** In this case we compare results for a similar environment in terms of mean den-

sity of galaxies but spatial correlation and cluster nature. The performance of PZWAV for the *Euclid* Mock catalogue with respect to the other shows less purity and less completeness, thus, as discussed in the previous case, the nature of the clusters and the spatial correlation of the field galaxies plays a key role in the cluster detection.

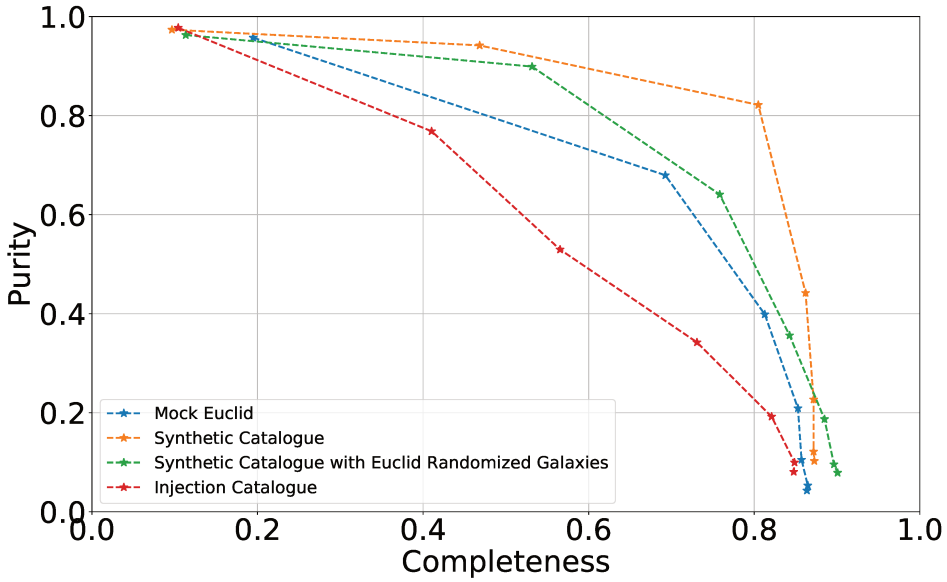


Figure 7.5 – Purity versus completeness for several cuts in detection's percentages, represented as dots, in increasing order from left to right, i.e., 1% of detections correspond to low completeness and high purity, and 99% of detections refers to high completeness and low purity. We chose a fixed value for $R_{MP}$ of 2 Mpc and a fixed value for $\Delta z$ of 0.2. The blue, orange, red and green curves correspond to the Mock catalogue, the synthetic catalogue, the injection catalogue, and the synthetic catalogue with *Euclid* 's randomized galaxies, respectively.

### 7.4.2 Richness Estimates

To compute the selection function we need the scaling relations between the mass and the direct observable of the clusters, that in our case, is the richness. As discussed before, PZWAV estimates the richness as a proxy of the amplitude of the peaks in the smooth galaxy density maps it generates.

Figure 7.6 shows the logarithm of the richness estimated from PZWAV versus the logarithm of the "true" richness coming from the cluster catalogues. The richness for the *Euclid* Mock catalogue, the synthetic catalogue, the injection catalogue and the synthetic catalogue with *Euclid*'s randomized galaxies is shown in red, blue, green and purple dots, respectively. We observe a significant difference between the estimation from the catalogues including only *Euclid* mock

clusters and the catalogues with pure synthetic clusters. For the latter, the richness is much lower. We observe vertical lines of the richness for the blue and purple dots. This is due to the fact that for the synthetic clusters those in the same bin in mass and redshift were simulated with the same number of galaxies. However, PZWAV gives a different estimation of the richness. Moreover, for the synthetic catalogue with *Euclid*'s randomized galaxies, the richness value is greater. Therefore, environment plays a key role in the richness estimation.

The richness for the injection catalogue is generally greater than for the synthetic catalogue. Nevertheless, the performance of PZWAV in terms of completeness and purity is worse for this case, as discussed previously. Our interpretation is that, the overlapped clusters are detected as a single cluster mixing the properties of the *Euclid* Mock and synthetic clusters. Thus, we would have less clusters detected (worse completeness) but large richnesses for the associated synthetic cluster.

Finally, we would like to come back to the fact that the synthetic clusters show overall lower richness with respect to *Euclid* Mock clusters. We do not understand this behaviour of PZWAV. One way to better understand this is to mimic the environment of the *Euclid* Mock field galaxies to disentangle between environmental effects and cluster nature effects.

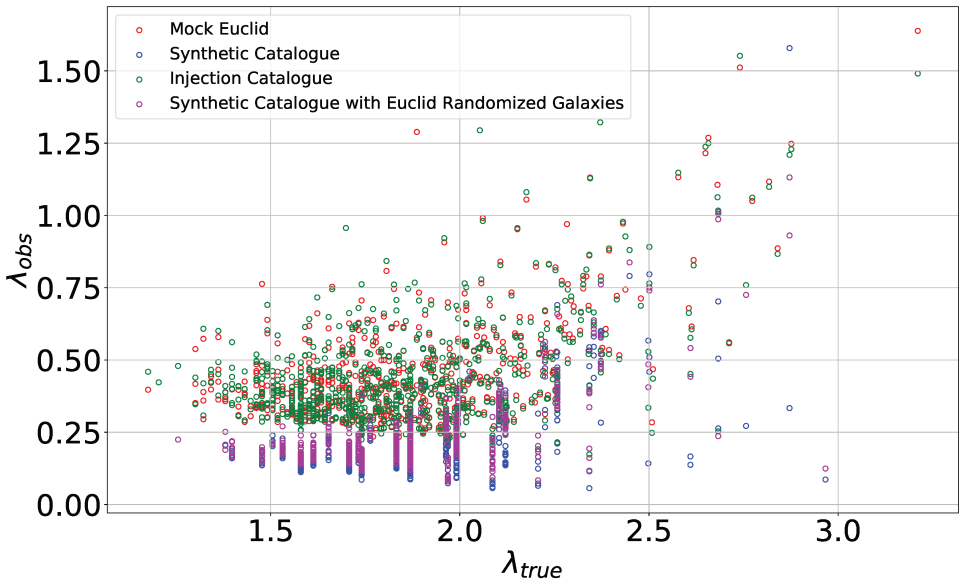


Figure 7.6 – Expected observed richness of the PZWAV detected clusters, $\lambda_{obs}$, versus the true richness, for $M^{mock}_{cluster} > 10^{14} M_{\odot}$ and $z^{mock}_{cluster} < 2$.

## 7.5 First Attempt to Estimate the Selection Function

For computing the SF we need to compute the probability of finding a cluster of a given true mass and redshift for a particular observational cut. Here, we perform a first attempt towards this final purpose.

We have started by selecting the 5% of detections with largest SNR, so that we have a large enough purity and completeness for the four used catalogues as represented in Figure 7.5. For each of those catalogues we have, first, estimated a simple log-linear scaling relation relating the true mass of the detected cluster to the PZWAV estimated richness, $\lambda_{obs}$. For computing this relation we perform a robust least-square fit. With this, we are able to convert the true masses of the simulated clusters (synthetic or *Euclid* Mock clusters) into an expected observed richness. Then, we compute the probability of finding a cluster of a given true mass or an equivalent extrapolated observed richness at a given redshift from the ratio of the number of detected clusters with respect to the total number of simulated clusters. This probability is shown in Figure 7.7 from top to bottom for the four catalogues considered: *Euclid* Mock catalogue, the synthetic catalogue, the injection catalogue, and the synthetic catalogue with *Euclid* 's Mock randomized galaxies.

First of all, we notice that the richness estimates, as discussed before, are very different between the different simulated catalogues. The expected observed richness for catalogues including *Euclid* Mock clusters shows lower dispersion that the ones including only synthetic clusters. Second, we observe that in general the probability of detecting a cluster is larger for catalogues for which the environment is featureless (synthetic catalogue and synthetic catalogue with *Euclid* 's Mock randomized galaxies) as we already observed previously. This may be just related to the fact that the richness estimates are very similar for this two catalogues. Furthermore, we find that the detection probabilities are consistent between these two catalogues. So we conclude that if the environment is featureless the results do not depend on properties of the field galaxies. Finally, The difference between the results for the *Euclid* Mock catalogue and the injection catalogue could be due to various aspects: (i) the difference in the properties in the clusters (synthetic vs *Euclid* Mock clusters), (ii) the fact that there is an overdensity of clusters in the injection catalogue and (iii) the difference of purity between the two cases.

This simplified analysis illustrates the complexity of selecting a methodology to estimate the selection function in a real case. We conclude that it is very important to be able to either reproduce or preserve the properties of the field galaxies in terms of the spatial correlation to ensure that the richness estimates in the true survey clusters and the simulated ones (those used for estimating the selection function) are consistent. This is true for both injection and full

simulation methods. We have also found that if using an injection method it is important to reproduce as accurately as possible the properties of the true survey clusters. Furthermore, in this case, it will be important to find a correct way to remove the detected true clusters prior to the injection of synthetic clusters.

The purpose of this analysis has been to better understand the difficulties in the construction of the selection function but more work is needed:

- Extend the analysis to a significant sky region (we have only considered 36 $\deg^2$).
- Understanding the differences in the richness estimates.
- Find a suitable way to correctly remove already detected clusters prior to injection.
- Introduce a realistic scaling relation for simulating the synthetic clusters, rather than using the true mock mass as we did.
- Investigate how to estimate the properties of real clusters: luminosity function and galaxy density distribution, which are needed to simulate the synthetic ones.
- Repeat the analysis using a different cluster finder (e.g., Amico [134, 133]).

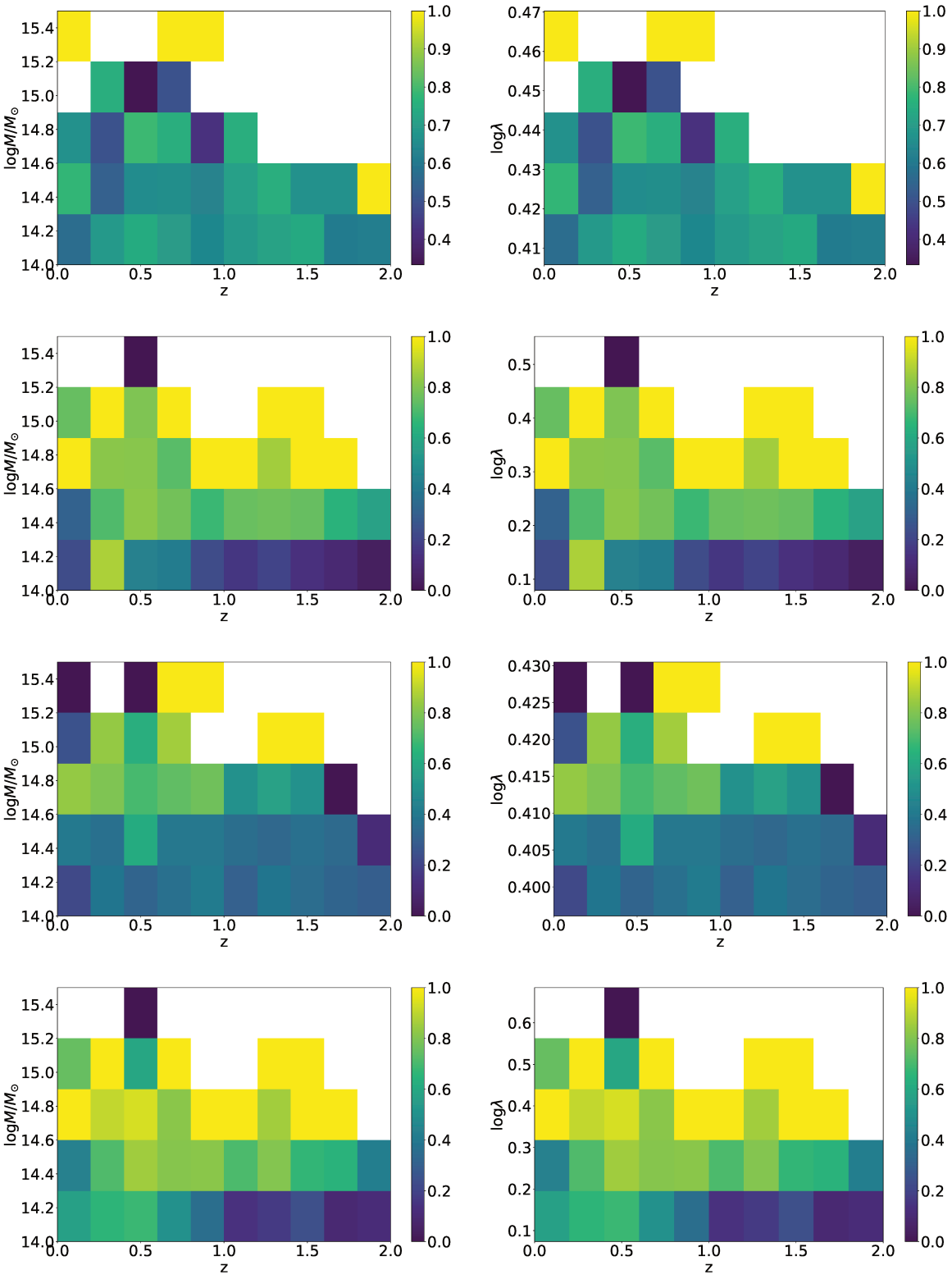


Figure 7.7 – Color coded probability of detecting a cluster of mass M (left column), or richness, $\lambda$ (right column), and redshift, z for a limit in true mass and redshift of $M^{mock}_{cluster} > 10^{14} M_{\odot}$ and $z^{mock}_{cluster} < 2$. From top to bottom we show the results for the four simulated catalogues considered: *Euclid* Mock catalogue, the synthetic catalogue, the injection catalogue, and the synthetic catalogue with *Euclid* 's Mock randomized galaxies.

## 7.6 Conclusions

In this chapter we have studied the perfomance of the cluster finder PZWAV for clusters more massives than $M > 10^{14} M_{\odot}$ and redshifts below $z < 2$ for four cases: the *Euclid* Mock catalogue, the synthetic catalogue, the injection catalogue and the synthetic catalogue with *Euclid* 's randomized galaxies.

We computed the dependency of the completeness and purity with the parameters $\Delta z$ and $\mathrm{R}_{MP}$ of the geometrical procedure to match the clusters between the catalogues. We conclude that $\Delta z$ is mainly affected by the environment while $\mathrm{R}_{MP}$ is more related to the cluster nature.

To compare the performance for the four catalogues we computed the completeness and purity for $\Delta z = 0.02$ and $\mathrm{R}_{MP} = 2$ Mpc, which we take as reference values. We observe better matching for the clusters simulated analytically. However, when the environment changes, the results are slightly different, worsen for a more complex environment which takes into account spatial correlations. The worst performance occurs with the injection catalogue. This can be due to the overlapping of the *Euclid* Mock clusters with the synthetic ones, which could produce overmergings and reduce the number of detected clusters. The best performance is for the synthetic catalogue with featureless environment and synthetic clusters. To explore more the PZWAV performance we would like to construct a catalogue that keeps both environment and spatial correlations, and adding more complex properties on the clusters.

Understanding variations on the PZWAV estimates of the richness of the detected clusters, which serves as a proxy of the mass, is essential for the determination of the selection function. We checked that for the synthetic clusters the richness estimates are significantly lower with respect to the catalogues that includes *Euclid* Mock clusters. However, the matching performance is worse for the latter. For this reason, it is important to make realistic simulations of galaxy clusters in order to improve the accuracy of the selection function.

Finally, in this chapter we have illustrated the complexity of constructing a selection function either using a mock catalogue or an injection catalogue methodology. We have shown that the properties of the field galaxies in terms of the spatial correlation plays a key role and need to be reproduce accurately in both methodologies. Further work is needed to better define how to go from the properties of the real survey detected clusters to a simulation of synthetic clusters.

Chapter 8

# The Three Hundred Project: towards a more realtistic synthetic catalog using the 300 cluster simulations



In this chapter we introduce the Three Hundred Project, a sample of 324 cluster regions constructed from hydrodynamical simulations. These regions have in their center a massive galaxy cluster, but other smaller clusters can be found around. Our main goal is to reproduce the analysis performed for the *Euclid* cluster galaxies' Mock catalogue in Chapter 5 by using more realistic cosmological simulations. We will study optical cluster properties: the luminosity function (LF) and the galaxy density radial distribution. We computed the LF for two different resolutions of the simulations, and we observed that at low resolution there is a significant lack of galaxies. This lead us to do an analysis of resolution effects in cluster properties for three different types of simulations: N-body at low and high resolution, and hydrodynamical at low resolution. We computed a resolution mass cut where the cluster properties are no longer

affected by resolution. The latter allows us to compute the galaxy density distribution, fit it by an analytical model, and study the evolution of its parameters with mass and redshift, as done in Chapter 5.

## 8.1 The Three Hundred Project

### 8.1.1 The 300th Cluster Catalogue

Large volumes are needed to find massive objects like galaxy clusters which entails the difficulty of modelling and simulating dark matter and baryonic physics with enough resolution for these large volumes. A solution to this are 'zoom' simulations, as adopted by the Three Hundred Project. For these large volumes are simulated by N-body dark-matter-only simulations, and only in the regions where a galaxy cluster is found, full-physics simulations are performed. The main disadvantage of 'zoom' simulations is that for having enough statistics it is necessary to run at least hundreds of independent simulations.

The N-body simulations for the Three Hundred Project are the MDPL2 MultiDark Simulations [189]. The latter constructs a 1 $h^{-1}$ Gpc cube containing $3840^3$ dark matter (DM) particles with a mass of 1.5 x $10^9$ $h^{-1}M_\odot$ each. Once the dark-matter-only simulations are performed, a cluster finder algorithm is ran. In this case the ROCKSTAR halo finder [190], which will look for dark matter haloes. A total of 324 spherical regions were extracted from the halo finder results, selecting as center for these regions the position of the most massive halo at redshift $z = 0$. The radius of each spherical region is 15 $h^{-1}$ Mpc that is much larger than the virial radius of the central cluster, which is the radius that encloses the mass that corresponds to approximately 98 times of the critical density of the Universe (at $z = 0$), as given by the Spherical Collapse model. The phase space initial conditions for the 324 selected regions are used to perform the 'zoom' re-simulations. For the study presented in this thesis, the 300th collaboration has ran these simulations in four different flavors based on the GADGET-X [191] code, which can generate both dark matter only or hydrodynamical simulations. The GADGET-X code is based on a modern version of the Smoothed Particle Hydrodynamics (SPH). The SPH is a numerical method for solving the fluid equations discretizing the continuous fluid in particles flowing in it, first proposed by [54, 55]. See Chapter 1 for more details. In the case of hydrodynamical simulations the GADGET-X code applies the following baryonic physical models:

- The gas treatment consists of an homogeneous UV background [192] and gas metal dependent cooling [193].
- The gas treatment consists of an homogeneous UV background [192] and gas metal dependent cooling [193].
- Star formation and stellar feedback are included using a stellar model by Tornatore et

al [194] and galactic stellar and substellar initial mass function by Chabrier et al [195]. Kinetic feedback [196], wind velocity of 350 km/s, thermal feedback and no gas mass loss are also implemented.

- Finally, GADGET-X includes black hole seeding and growth and active galactic nuclei (AGN) feedback [197].

The four different flavours in the re-simulated regions are:

1. ***LR DMONLY:*** Dark matter only simulations with a dark matter particle resolution of 1.5 x $10^9$ $h^{-1}M_\odot$.
2. ***HR DMONLY:*** Dark matter only simulations at high resolution. With respect to *LR DMONLY*, it has twice particles per dimension, thus, $7680^3$, with eight times less mass per particle i.e., 1.8 x $10^8$ $h^{-1}$ $M_\odot$ each. For these simulations we only have 68 regions, not 324 as for the low resolution ones. Hence, when comparing with the low resolution simulations with the *HR DMONLY* ones, we will use the 68 regions for which we have high resolution data.
3. ***LR HYDRO:*** Full-physics hydrodynamics zoom simulations at low resolution. These simulations are at the same resolution as the *LR DMONLY* simulations.
4. ***HR HYDRO:*** Full-physics hydrodynamics zoom simulations at high resolution. These simulations are at the same resolution as the *HR DMONLY* simulations. For these simulations we only have 1 region, due to the high computational cost.

For each simulation, a total of 128 snapshots in redshift are stored, for a redshift range between $z = 17$ and $z = 0$. As explained in Chapter 5, the N-body simulations for the *Euclid* Mock Catalogue have a resolution of $5200^3$ particles for a 1 $h^{-1}$ Gpc box. That means that the low resolution simulations for the Three Hundred project and the high resolution ones are below and above the resolution of *Euclid*, respectively.

Once the regions are re-simulated, they are analysed by the Amiga's Halo Finder (AHF) [198], producing a catalogue with haloes found within the regions. In the case of hydrodynamical simulations, for each halo different properties are computed, such as its radius $R_{200}$, mass $M_{200}$, density profile, galaxy luminosities for several spectral bands covering from far-UV to radio. The galaxy luminosities are computed from the identified stars of the AHF finder using the STARDUST code [199]. The spectral energy distribution (SED) of each galaxy is convolved with the bandpass of each photometric filter to compute the galaxy luminosity. In the case of dark matter only simulations we have the same properties except those related to baryon physics (e.g., stellar mass and luminosities). Each halo can have smaller haloes gravitationally bounded to it, which we will call subhaloes, with their own properties. The more massive and central

halo is known as the host halo. Then the final data set is composed of 324 regions, for the low resolution cases, 68 regions for the high resolution dark matter only and 1 region for the high resolution hydrodynamical simulations. We have a low mass threshold for the central halo of $6 \cdot 10^{14} h^{-1} M_{\odot}$ at $z = 0$. Other haloes, which we do not consider, can be found out of the sphere of the central halo, in the 15 $h^{-1}$Mpc region.

### 8.1.2 Identifying Galaxy Cluster Members in the 300th Cluster Catalogue.

In this section we explain how to construct the dataset that will be used for all the analysis during this chapter. As discussed, we have access to three different cluster catalogues corresponding to the simulations describes before: *LR DMONLY*, *HR DMONLY* and the low resolution hydrodynamical simulations or *LR HYDRO*, that shares the same resolution with the low resolution dark-matter-only simulations. For the *HR HYDRO* we only have one single re-simulated region.

For our dataset only the most massive and central cluster of the 15 $h^{-1}$ Mpc region, also called host halo, is considered. This is done for ensuring large enough clusters that can host a significant number subhalos or substructures inside, implying higher statistics. A subhalo is a gravitationally bound structure to the host halo. It can be a small cluster with substructures bound to it, or it can also be a single galaxy bound directly to the host halo. Here, as we will explain later, a subhalo that is identified as a cluster will be split into its various galaxies, and we will consider them as bounded directly to the host halo. However, there is no clear definition for what a galaxy is in the 300th catalogue, i.e., we only have halo properties, that can be a galaxy, or another type of object. Therefore, we need to find first a definition for galaxy, identify them and create a cluster's galaxy members catalogue. Once this is done, in the following, a galaxy can be called substructure or subhalo and viceversa.

To define a galaxy, first we consider a mass threshold, for the subhalo mass, for each of three simulations. This translates into a mass resolution limit. Notice that we consider the particle mass instead of the number of particles as our threshold because for the different simulations the particle mass is different. Thus, the same number of particles does not translate into the same subhalo mass. Taking into account the particle mass and the simulation resolution we adopt the following cuts:

- For *LR DMONLY*, we consider 4 x $10^{10}$ $M_{\odot}$ as the lower limit. This is equivalent to considering that the substructure is formed by, at least, 20 dark matter particles.
- For *HR DMONLY*, considering the same mass threshold as for the low resolution ones, it leads to be at least 160 dark matter particles, because the resolution is 8 times higher. Since the resolution is significantly higher, we can vary this threshold with respect to

the low resolution ones. So to check resolution effects, we choose a mass threshold of 1 x $10^{10}M_{\odot}$, corresponding to 50 dark matter particles.

- The *LR HYDRO*) share resolution with the *LR DMONLY* case. However the particle mass is different because in this case we have dark matter particles and gas particles. This means that the same mass threshold between *LR DMONLY* and *LR HYDRO* does not translate into the same number of particles. A hydrodynamical simulated structure can have the same mass as a dark-matter-only one, but without dark matter particles. We have to ensure that a galaxy has dark matter particles. For this reason we apply, on top of the mass threshold, a number of particles threshold. Thus we consider a mass threshold of 4 x $10^{10}$ $M_{\odot}$ (for *LR DMONLY* case) and a number of particles threshold of 10. We do not put more than 20 particles, because we want the same mass threshold for simulations that share the same resolution. Doing this, we assure the presence of dark matter components in the baryonic structures, and also a minimum mass for having enough resolution.

The next step is to distinguish between substructures that are directly bounded to the host halo, and those that are inside a subhalo, called subsubtructures. The AHF cluster detection algorithm defines haloes from a maxima in the density field. It is possible to find some group of galaxies inside a cluster that the detection algorithm interpret as another halo. For our purposes, these galaxies will belong to the host halo. However, in the catalogue they appear as secondary galaxies bounded to secondary haloes, so we have to make some changes. The first thing is to consider that a subhalo is formed by a brightest central galaxy (BCG) and subsubstructures. In the catalogue, the available information for a halo (either host halo or subhalo) is the bound mass, i.e., the sum of the virial masses of every structure, including the BCG, that is directly linked to the halo. Then, for finding the mass of the BCG it is necessary to substract the sum of all the masses of the substructures to the total bound mass. Once this is done, a subhalo then is treated as a group of galaxies, including a BCG, that can be linked directly to the host halo, and then they are not considered subsubstructures anymore. After this process, our dataset is formed by a list of host halos with galaxies linked to them. In the case of the hydrodynamical simulations, we need to consider the mass ratio between the dark matter and the star content of the galaxy. We assume that the mass of a real galaxy is mainly coming from the dark matter halo surrounding the stars and gas, so we impose that the stellar mass component is not higher than 30% of the total mass.

The last part of the clean data process and galaxy selection, is to avoid contamination from low resolution particles that initially were outside the region of interest. To get rid of these particles and maintain only well resolved structures, we indicate the fraction of mass in the high resolution particles. We have considered a minimum mass fraction resolution of 0.999 (where

the perfect zero contamination level would be one).

Finally we consider three cluster catalogues, with the 3D properties of haloes and subhaloes. We have 324 regions for the two low resolution simulations, *LR DMONLY* and *LR HYDRO*, and 68 regions for the high resolution ones, *HR DMONLY*. We have a host halo per region with a BCG and bounded galaxies, all of them free of contamination. As discussed above, for being able to compare the three simulations, we will choose the same 68 regions.

## 8.2 Resolution Effects

The primary goal of this chapter was to redo the work done in Chapter 5 but using more realistic simulations. Therefore, we want to compute the galaxy density radial distribution and the LF for the 300th clusters for several bins in mass and redshift. Later, we want to fit these properties by analytical methods to finally construct a synthetic catalogue. We expect from this catalogue more realistic cluster properties because it comes from more realistic simulations. We repeat the process explained in Chapter 5.

### 8.2.1 Resolution Effects on the Luminosity Function

First, we study is the LF, defined as the number of galaxies divided by the magnitude bin difference and the total cluster area, i.e., $A = \sum \pi \, R^2_{i,200}$, with i for each cluster in the bin. The LF can only be computed for simulations with baryonic information, because we need the magnitude of galaxies. Thus, we compute the LF for the low resolution hydrodynamical simulations, *LR HYDRO*, which contains 324 clusters. To compare with the results of SAMs corresponding to the *Euclid* Mock catalogue, we establish a magnitude upper limit of 24. Figure 8.1 shows the luminosity function at redshift $z = 1$, for the 324 central clusters. The black points with uncertainties are the number of galaxies per magnitude bin difference per total clusters' area with a Poisson distributed uncertainty over the number of galaxies (see Chapter 5). The LF is fitted by a Schechter model, following Equation 5.5. We use a MCMC to perform the fit, as explained in Chapter 5. The red and green shaded areas are the $1\sigma$ and $2\sigma$ uncertainties over the best-fit, respectively. As we observe in the figure, the number of galaxies drop after an apparent magnitude of 19, and the model does not fit the data, as one would expect from the shape for the large magnitudes.

The drop in the number of faint galaxies could be caused because of a lack of resolution in the hydrodynamical simulations. To test this hypothesis, the 300th Collaboration has run a hydrodynamical simulation of one of the 324 cluster regions at high resolution. We present in Figure 8.2 the LF for the high resolution hydrodynamical simulation. The process to compute

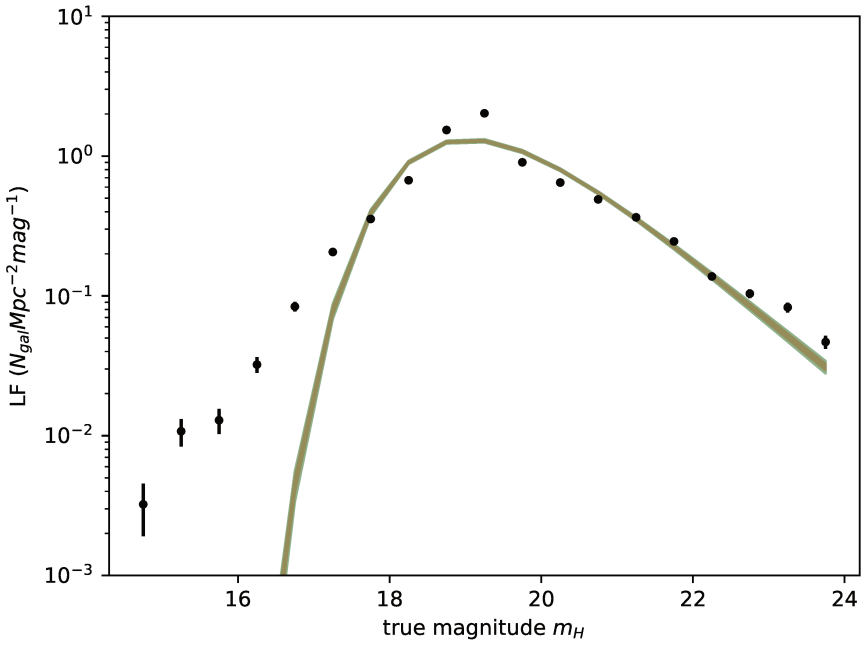


Figure 8.1 – Luminosity function fitted by a Schechter model. The 324 low resolution hydrodynamical clusters at z = 1. The black points with the error bars are the number of galaxies per magnitude bin difference per total clusters' area with a Poisson distributed uncertainty over the number of galaxies. The red and green shaded areas are the $1\sigma$ and $2\sigma$ uncertainties for the best-fit, respectively.

the luminosity function as well as the model fitting is the same as for the low resolution simulations. In this case, the drop of faint galaxies observed in Figure 8.1 does not appear for the high resolution hydrodynamical simulation and the Schechter model is a good fit to the data. We can therefore conclude that increasing the resolution for hydrodynamical simulations solves the drop in the faint end of the low resolution simulations. Thus, we also conclude that it must be possible to model analytically the LF for the 300th cluster catalogue, and compare the results with the *Euclid* semianalytical simulations. However, we only have one cluster at high resolution so we do not have enough statistics to perform a complete analysis. The computational cost of hydrodynamical simulations so high that has not been possible to go any further on this during this thesis.

For this we will study the galaxy (subhalo) mass function and compare between low and high resolution simulations. The fact of having found that the lack of resolution can affect significantly the results, leads us to start an analysis of resolution effects before considering the study of the galaxy density radial distribution in the cluster.

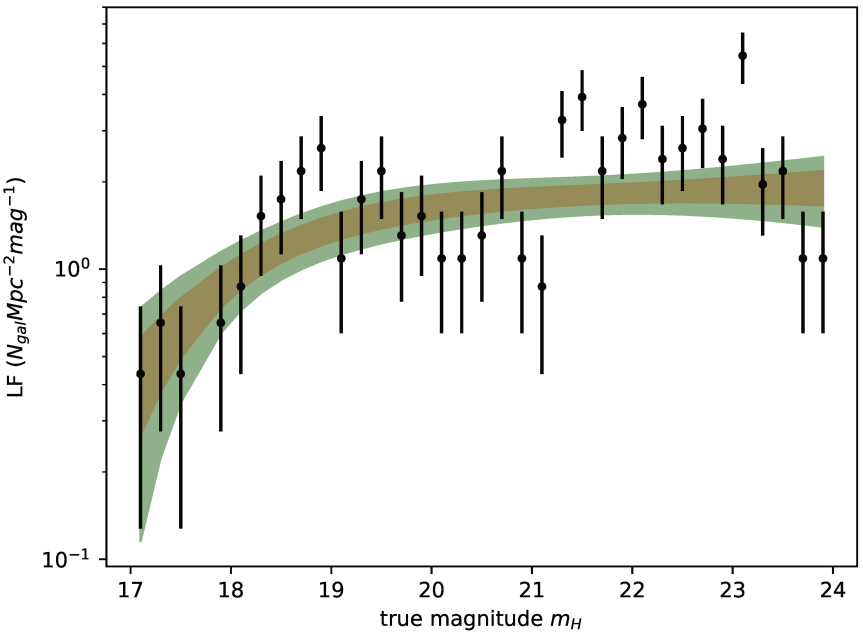


Figure 8.2 – Luminosity function fitted to a Schechter model, for 1 high resolution hydrodynamical cluster at z = 1. The black points with the error bars are the number of galaxies per magnitude bin difference per total clusters' area with a Poisson distributed uncertainty over the number of galaxies. The blue dashed line and the red and green shaded areas are the best-fit and $1\sigma$ and $2\sigma$ spread, respectively.

### 8.2.2 Galaxy (subhalo) Mass Function

The halo mass function gives the number of halo member galaxies of a given mass, relative to the mass of the halo. The final mass resolution of the simulation will imprint in the halo mass function as a deficit of low relative mass subhalos. We start by comparing the low resolution and high resolution dark matter only simulations. In practice, we study the cumulative galaxy mass function defined as the total number of cluster member galaxies' with the relative mass (ratio between their mass and their host halo mass) below a given value.

As an example, we present in Figure 8.3, the 3D cumulative galaxy mass function at redshift $z = 0$ for the three mass bins presented in Table 8.1: *Bin 1*, *Bin 2* and *Bin 3* from top to bottom, respectively. Each cluster mass bin, was chosen based on the cluster mass range of the 68 regions. We have selected as narrow bins in mass as possible to ensure that cluster properties

| Mass Bin Names | Cluster Mass |
|---|---|
| *Bin 1* | 7 x $10^{14}$ $M_\odot$ < M < 1 x $10^{15}$ $M_\odot$ |
| *Bin 2* | 1 x $10^{15}$ $M_\odot$ < M < 2 x $10^{15}$ $M_\odot$ |
| *Bin 3* | 2 x $10^{15}$ $M_\odot$ < M < 1 x $10^{16}$ $M_\odot$ |

Table 8.1 – Cluster mass bins considered in this chapter. The first bin, called *Bin 1*, corresponds to cluster masses: 7 x $10^{14}$ $M_\odot$ - 1 x $10^{15}$ $M_\odot$. The second bin, called *Bin 2*, corresponds to cluster masses: 1 x $10^{15}$ $M_\odot$ - 2 x $10^{15}$ $M_\odot$. The third bin, called *Bin 3*, corresponds to cluster masses: 2 x $10^{15}$ $M_\odot$ - 1 x $10^{16}$ $M_\odot$.

within a bin are similar while having a sufficient number of clusters per bin. The color code used in this figure will be the same for the next results during the chapter

- High resolution dark matter only simulations, *HR DMONLY*, in red.
- Low resolution hydrodynamical simulations, *LR HYDRO*, in black.
- Low resolution dark matter only simulations, *LR DMONLY*, in blue.

In the Figure, each line corresponds to the mean value of the 3D galaxy mass functions of the clusters in the bin. The shaded regions are the $1\sigma$ uncertainties coming from the standard deviation across clusters. We compute the cumulative function, i.e., when increasing the mass threshold, we include galaxies which masses are above the threshold. For this reason we can observe larger uncertainties at high substructure mass because we have less remaining objects in the cluster.

First we compare the two different resolution dark matter only simulations. The galaxy mass function for both resolutions converge to the same number of substructures for the most massive ones. The main difference between the two is seen for the smaller substructures. We find many more of them for the high resolution simulations. For both mass functions we observe that the cumulative number of structures is constant for the smallest masses, indicating that there are no structures smaller than a certain threshold in relative mass ($M_{substructure}/M_{parent}$). This threshold is different for the low and high resolution simulations. The latter show many more galaxies and extend to smaller masses. This feature is common for the three bins in mass. The only difference is that when the cluster mass is higher, we find more substructures, as expected. We can conclude from this figure that the lack of resolution affects only the formation of smaller structures, while conserving the total mass of the cluster. Galaxies that are bigger size in the low resolution are divided in smaller ones for the high resolution.

We check now how the resolution effects affect the low resolution hydrodynamical simulations, *LR HYDRO*. For this we consider the same 68 clusters that we have used for the dark

matter only simulations and we compute the 3D cumulative galaxy mass function on the *LR HYDRO* simulations, in the same three cluster mass bins. In Figure 8.3, we show the results for the hydrodynamical simulations in black, at a redshift z = 0. We can see that the three simulations converge for the more massive substructures, and again the main differences come from the low relative mass galaxies.

The slope for the *LR HYDRO* simulations is closer to the *HR DMONLY* but still we can see that the black line is slightly above the red one in the low mass regime. For the same resolution, i.e., the blue and black lines, we find more objects for the hydro simulation, and these objects are less massive. This could be due to the fact that baryonic physics avoid the strip out of the particles because of gas cooling process.

For the hydrodynamical simulation we can see in general more objects than for the *LR DMONLY*, but less than for *HR DMONLY*. This behaviour is the same for every cluster mass bin. We also find, as expected, more substructures when the mass of the clusters are larger. We conclude that baryonic physics diminish the ripping out of the objects because of cooling down processes of the gas, so we keep more smaller substructures in comparison to the same resolution dark-matter-only simulations. Also, increasing the resolution in dark matter only simulations allows us to get a closer distribution to the hydrodynamical simulations even though we still find differences for both simulations. This means that improving the resolution in dark-matter-only simulations is not equivalent to adding baryonic physics in the simulations.

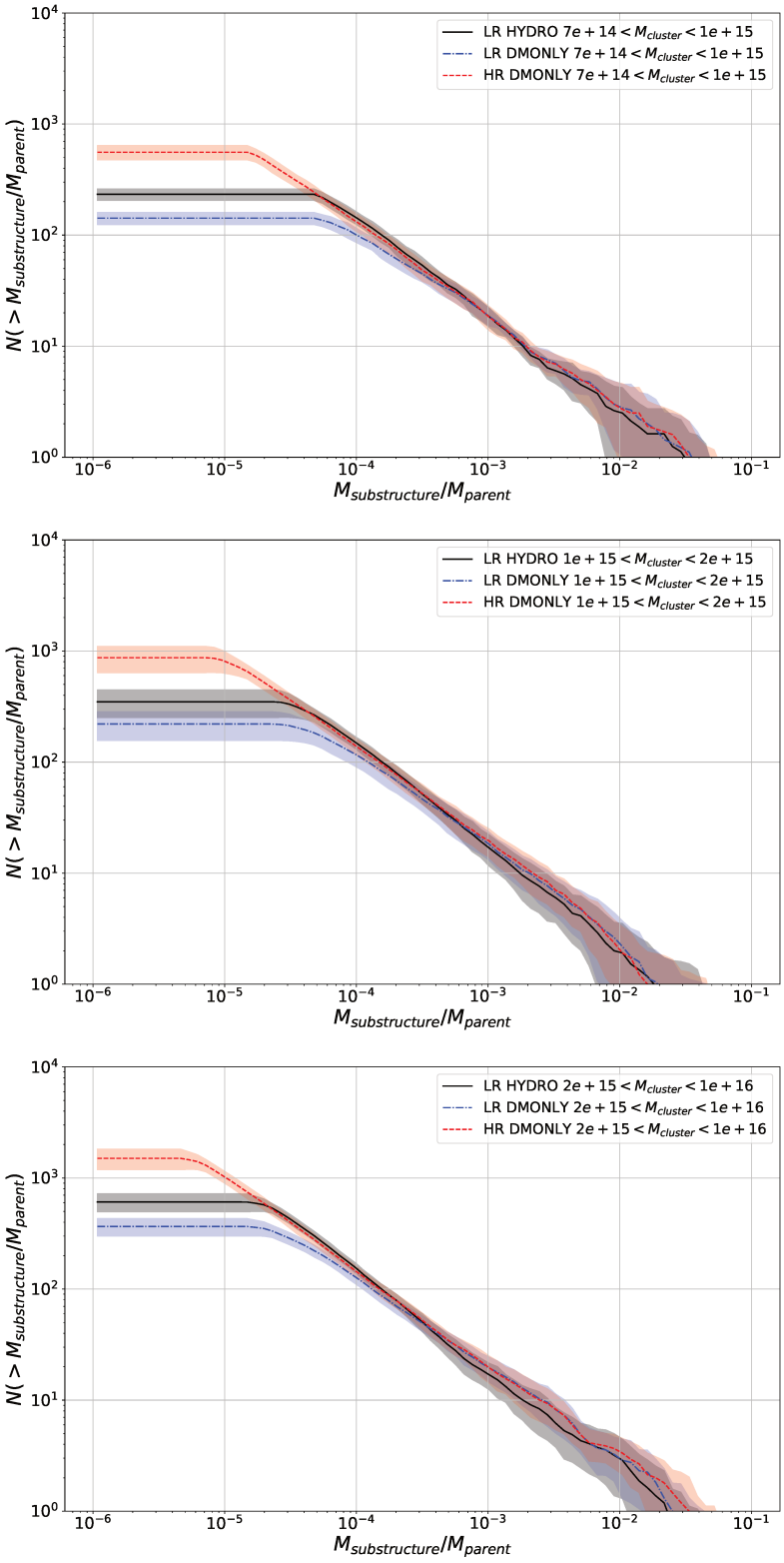


Figure 8.3 – 3D cumulative galaxy mass function in blue, red and black, for low and high resolution dark-matter-only and hydrodynamical simulations, respectively. From top to bottom the three different cluster mass bins: *Bin 1*, *Bin 2* and *Bin 3*, respectively, as presented in Table 8.1. Each line corresponds to the mean value of 3D galaxy mass functions of the clusters in the bin. The shaded regions are the $1\sigma$ uncertainties coming from the standard deviation across clusters.

### 8.2.3 Galaxy Mass Function Variation with Mass and Redshift

For the previous results we concentrated on redshift $z = 0$. However the redshift evolution of the galaxy mass function is fundamental for understanding structure formation processes. We present in Figure 8.4 the redshift evolution from $z = 0$ to $z = 1$ of the 3D cumulative galaxy mass function for two different cluster mass bins: *Bin 1* and *Bin 3*, following Table 8.1, and for the *LR HYDRO*, *LR DMONLY* and *HR DMONLY* simulations.

The color coding that will be used in plots for the redshift evolution during the chapter will be

- Redshift z = 0 in blue.
- Redshift z = 0.3 in orange.
- Redshift z = 0.5 in green.
- Redshift z = 0.8 in red.
- Redshift z = 1 in purple.

The upper row of the figure represents the *Bin 1* cluster mass ($7 \times 10^{14}$ $M_\odot$ $< M <$ $1 \times 10^{15}$ $M_\odot$) for *LR HYDRO*, *LR DMONLY* and *HR DMONLY* simulations from left to right, respectively. Each line corresponds to the mean value of 3D galaxy mass functions of the clusters in the bin. The shaded regions are the $1\sigma$ uncertainties coming from the standard deviation across clusters. For each simulation the number of substructures as a function of redshift is almost constant. This would be consistent with the self similar scenario for cluster formation. Clusters are a copy one of the others, that means that if they have the same mass, they have the same size, and their formation history is the same. Nevertheless, there is an evolution of the number of galaxies with redshift, so the clusters are not exactly self-similar one to the other, finding more galaxies at higher redshift, even though this difference is inside the $1\sigma$ uncertainties. We observe the same behaviour for the *Bin 2* in mass in the middle row of the figure.

The bottom row of the figure shows the results for the *Bin 3* in mass. For $z > 0.5$ we have no clusters in this mass range. For $z = 0.5$ we find only one cluster and therefore the uncertainties, computed from the dispersion among clusters, are undefined. The feature of the 3D galaxy mass function evolution with redshift is the same for both bins in mass. However, the only difference is that at higher cluster mass, there are more galaxies, as expected.

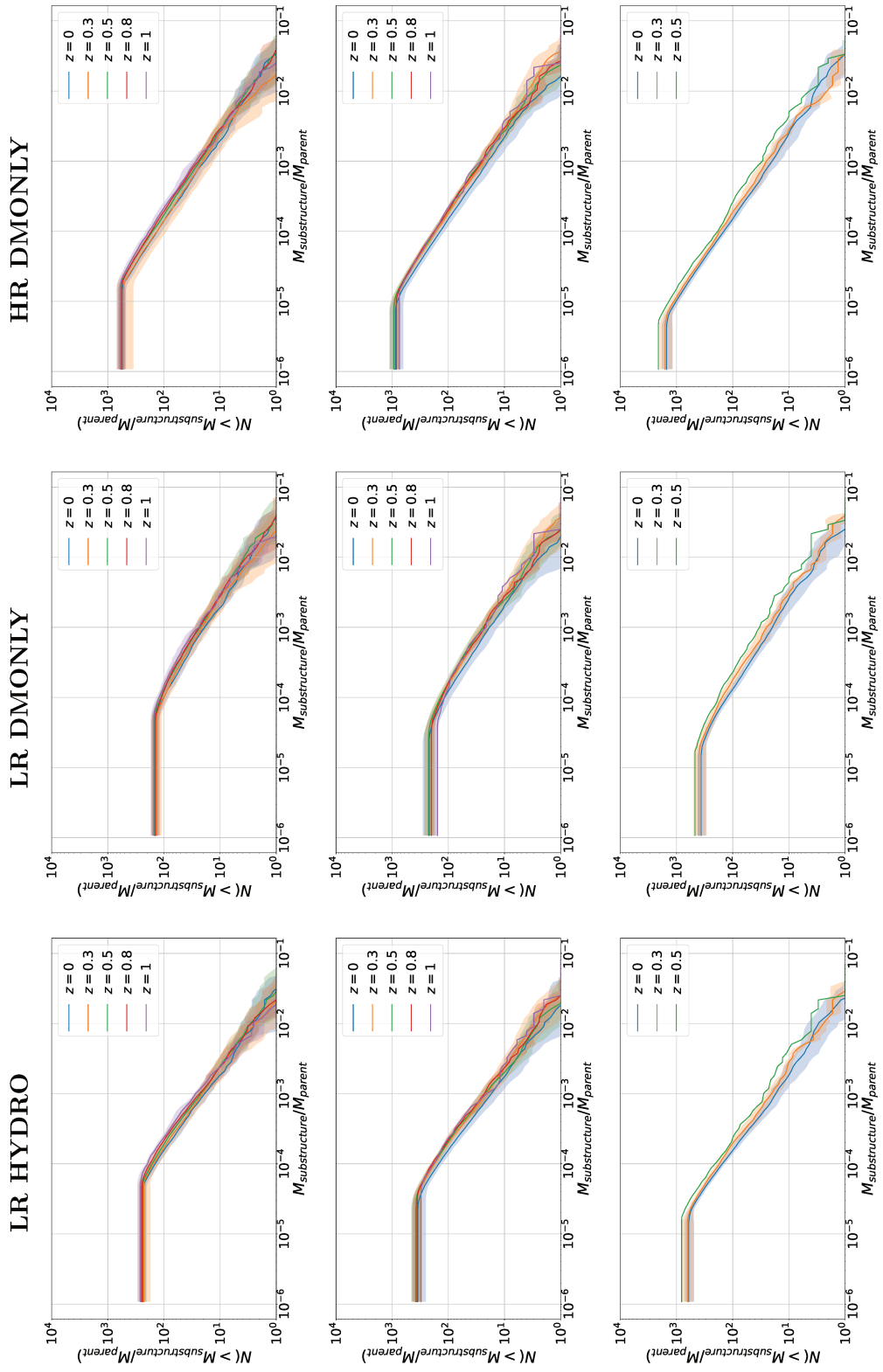


Figure 8.4 – 3D cumulative galaxy mass function redshift evolution from z = 0 to z = 1 for it LR HYDRO, *LR DMONLY* and *HR DMONLY* from left to right, respectively. From top to bottom the three cluster mass bins: *Bin 1*, *Bin 2* and *Bin 3* as explained in Table 8.1. For *Bin 3* we have no clusters of z > 0.5. Each line corresponds to the mean value of 3D galaxy mass functions of the clusters in the bin. The shaded regions are the 1$\sigma$ dispersion coming from the standard deviation across clusters.

### 8.2.4 Modelling and Fitting of the the Galaxy Mass Function

The 3D cumulative galaxy mass function can be well approximated by a power-law function as in [200],

$$N_m = N_{-4}\left(\frac{m_{sub}/M_{vir}}{M_\odot}\right)^{\alpha}, \tag{8.1}$$

where $N_{-4}$ is a normalization, $\alpha$ the slope and $m_{sub}/M_{vir}$ is the ratio between the virial mass of the substructures (galaxies) and their host halo. For the best-fit value parameters we perform a least square fits of the mean value accounting for the uncertainties computed from the dispersion across clusters as shown in Figure 8.3. In addition, we also perform a fit of the cumulative galaxy mass function per clusters and we compute the dispersion of the best-fit parameters across clusters.

We present in Figure 8.5 the evolution with redshift for both, the normalization and the slope parameters of the power-law fit, for the three cluster mass bins discussed above, and for the three types of simulations we are considering. From left to right we represent the results for the three cluster mass bins *Bin 1*, *Bin 2* and *Bin 3*, respectively, as defined in Table 8.1. As in the previous plots, the black color represents *LR HYDRO*, the red one *HR DMONLY* and the blue *LR DMONLY*. The redshift interval we are considering is $0 < z < 1$. The points represents the best fit parameters for the power-law fit over the mean 3D cumulative galaxy mass function (distribution shown in Figure 8.3). The uncertainties are computed from the dispersion of the power-law fit of the 3D cumulative galaxy mass function distribution per cluster, and from the intrinsic uncertainty of the fit of the mean cumulative galaxy mass function. The upper row of the figure corresponds to the slope parameter, $\alpha$. We see that we have a larger slope, in absolute value, for the *LR HYDRO*, followed by *HR DMONLY* and, finally, *LR DMONLY*, as we can check also in the Figure 8.3. In terms of redshift, the evolution is small, almost non-existent, and as expected from Figure 8.4. For the more massive bin, the plot in the right panel, we can see, that for $z > 0.5$, as we have seen in the Figure 8.4, we do not find clusters, so we show no results. In addition, as we discussed for Figure 8.4, at $z = 0.5$ the uncertainties are small because we only have one cluster. For the normalization parameter, $N_{-4}$, we find equivalent results, with no evolution with redshift. These results are in agreement with Dolag et al. [200], which did a similar analysis but in a smaller cluster sample both in mass and redshift.

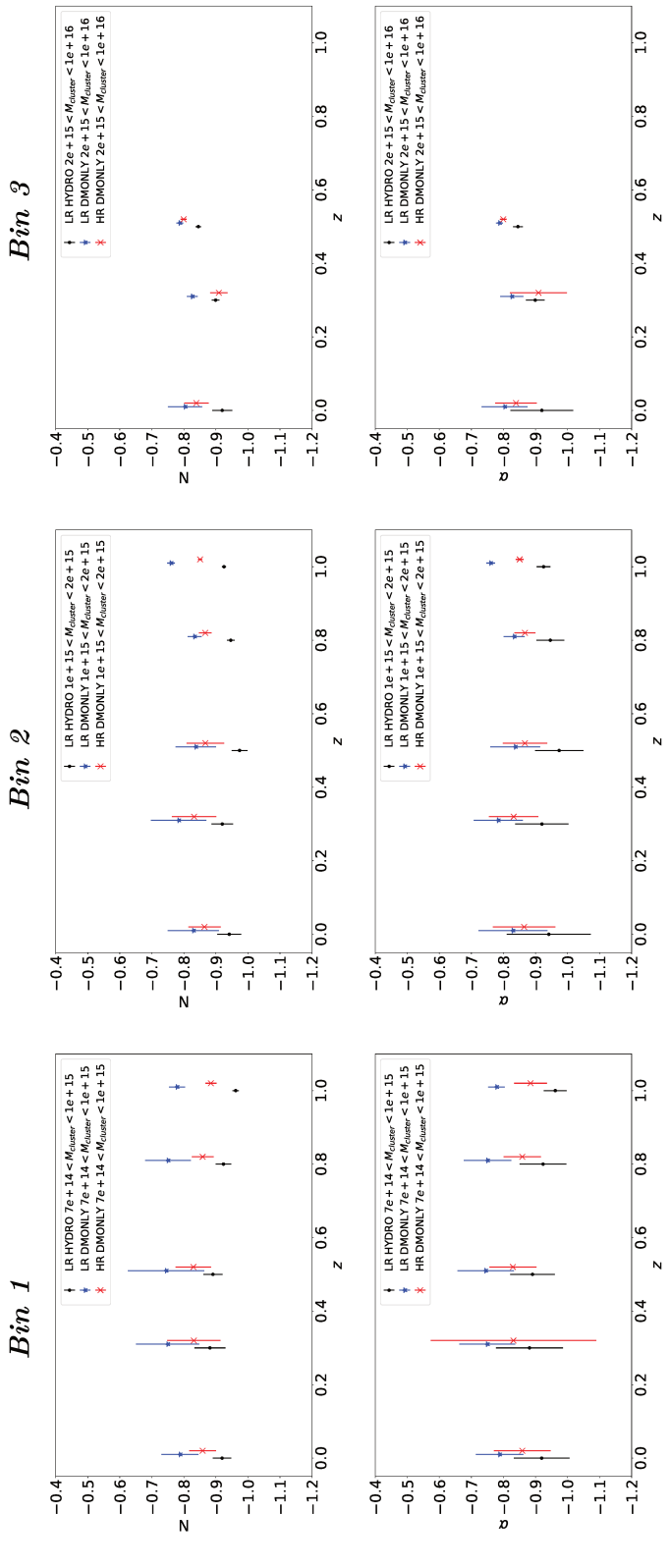


Figure 8.5 – Redshift and mass evolution of the parameters from the power-law fit of the 3D cumulative galaxy mass function. The upper row represents the $\alpha$ parameter and the bottom row the normalization $N_{-4}$, for $z = 0$ up to $z = 1$. From left to right we present results for the different cluster mass bins of Table 8.1: *Bin 1*, *Bin 2* and *Bin 3*, respectively. The results are shifted with respect to the *LR HYDRO* for visualization purposes.

### 8.2.5 Resolution Cuts for the Galaxy Mass Function

For concluding this analysis, we are going to establish galaxy mass cuts in order to be able to compare the different simulations in a region where resolution effects do not bias the recovered cluster properties. We present in Figure 8.6 for z = 0 and for the least massive cluster mass bin, *Bin 1*, the ratio between the 3D cumulative galaxy mass function for every simulation with respect of *HR DMONLY* one. We consider the latter as a reference model as it is less affected by resolution effects. In the figure, the black line is the ratio between *LR HYDRO* and *HR DMONLY*. And the blue line corresponds to the ratio of the two dark matter only simulations, *LR DMONLY* and *HR DMONLY*. The vertical lines in the figure represent the mass cut we choose, and the color is equivalent to the color of the curve they refer to. We consider that resolution effects are negligible when the ratio is approximately one, just before decreasing.

In the case of the hydrodynamical simulation, we can see an excess of galaxies between the mass ratio cut ($4 \times 10^{-5}$) and the one for low resolution ($4 \times 10^{-4}$). This is also appreciated in Figure 8.3, where even if the curves between the high resolution simulation and the hydrodynamical one are overlapping we can see that the black curve is a slightly above the red one for the low mass galaxy range. As we explained, this is probably due to the inclusion of baryon physics in the simulations. Due to cooling down of gas, the smallest galaxies are not ripped out. These two mass cuts will be used in the following, when comparing the properties of clusters in the three types of simulations considered. If we want to use the low resolution dark matter only simulation we have to use its resolution mass cut in the other simulations, for avoiding differences in the physics results related just to the resolution. Likewise, when using the hydrodynamical simulations, we will use its resolution mass cut. Here we only present, as an example, the mass cut for a single redshift and a mass bin, but we have computed it for every redshift and mass bin used before. The mass cuts will be named from now on as

- *No Cut*, if there is no resolution mass cut.
- *HYDRO Cut*, for the *LR HYDRO* resolution mass cut.
- *LR Cut*, for the *LR DMONLY* resolution mass cut.

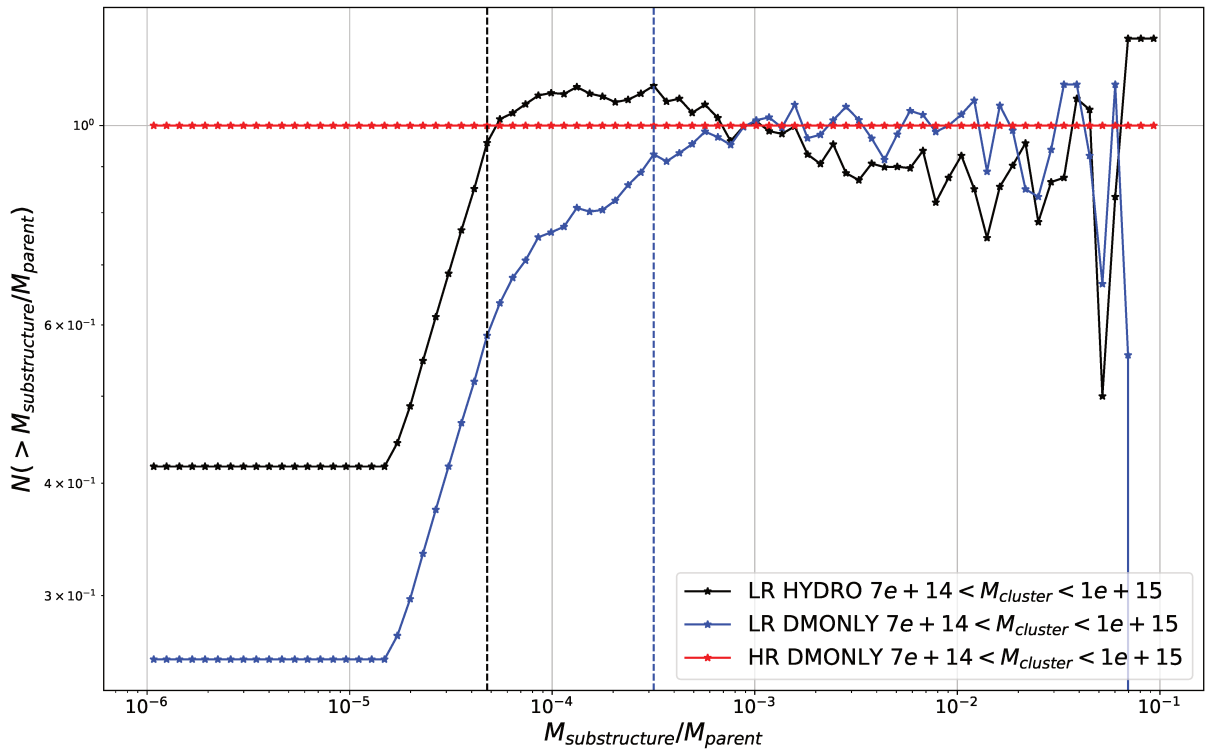


Figure 8.6 – Ratio between the 3D cumulative galaxy mass functions for the three simulations, taking *HR DMONLY* as a fiducial model, for cluster masses in the range 7 x $10^{14}$ $M_\odot$ < M < 1 x $10^{15}$ $M_\odot$ and at $z = 0$.

## 8.3 Galaxy Density Distribution

### 8.3.1 Methodology

Now that we have checked the resolution limits, for each simulation we compute the galaxy density profiles, which represents the number of galaxies in a spherical volume at a distance $r$ from the cluster center in $R_{200}$ units, as we did for *Euclid* Cluster Galaxies' Mock catalogue in Chapter 5.

We use the same bins in mass and redshift for computing the galaxy density profiles, as the ones in the previous sections. Figure 8.7 shows the 3D cumulative galaxy density profiles at redshift z = 0, for the three mass bins, from top to bottom, *Bin 1*, *Bin 2*, *Bin 3* (see Table 8.1), and for two resolution mass cuts, discussed before. In the figures in the left column, where we have applied the *HYDRO Cut*, we compare *LR HYDRO* with *HR DMONLY*. The right column of the figure, where we have applied the *LR Cut*, we compare the three types of simulations *LR HYDRO*, *LR DMONLY* and *HR DMONLY*. The data points are shifted with respect to *LR HYDRO* for visualization purposes.

To compute the galaxy density profiles we use equally spaced logarithmic bins and the cumulative distribution. For a spherical shell at a distance $r$ from the cluster center, we account for all the galaxies that are inside this sphere, and we compute the volume as $4\pi r^3/3$. The data points correspond to the mean value of the number of galaxies across all the clusters in the mass bin and over the volume of the bin. The uncertainties are the $1\sigma$ dispersion over the mean, as done in Chapter 5.

When using the *HYDRO Cut* (left column of Figure 8.7), the *LR HYDRO* simulation shows larger density towards the center of the cluster for all the bins in mass, and this difference increases with the mass bin. In terms of the mass bins, the galaxy density is higher when the mass is higher. Nevertheless, both profiles converge on the outskirts. Taking into account the results in Section 8.2, this means that the smaller objects that survive on the hydrodynamical simulations, are located towards the center. The fact that this increases with the cluster mass can be due to the fact that for a bigger size cluster there can be more fragmented galaxies, i.e., massive galaxies divided in smaller ones, that can be seen in the *LR HYDRO* simulations.

For the *LR Cut*, right column of the figure, the three types of simulations converge within $1\sigma$. This is expected because the small structures, which are the discrepancies between the hydrodynamical and dark matter only simulations, are not kept. Nevertheless, the slope for the *LR HYDRO* is higher as well as the mean value of the galaxy density.

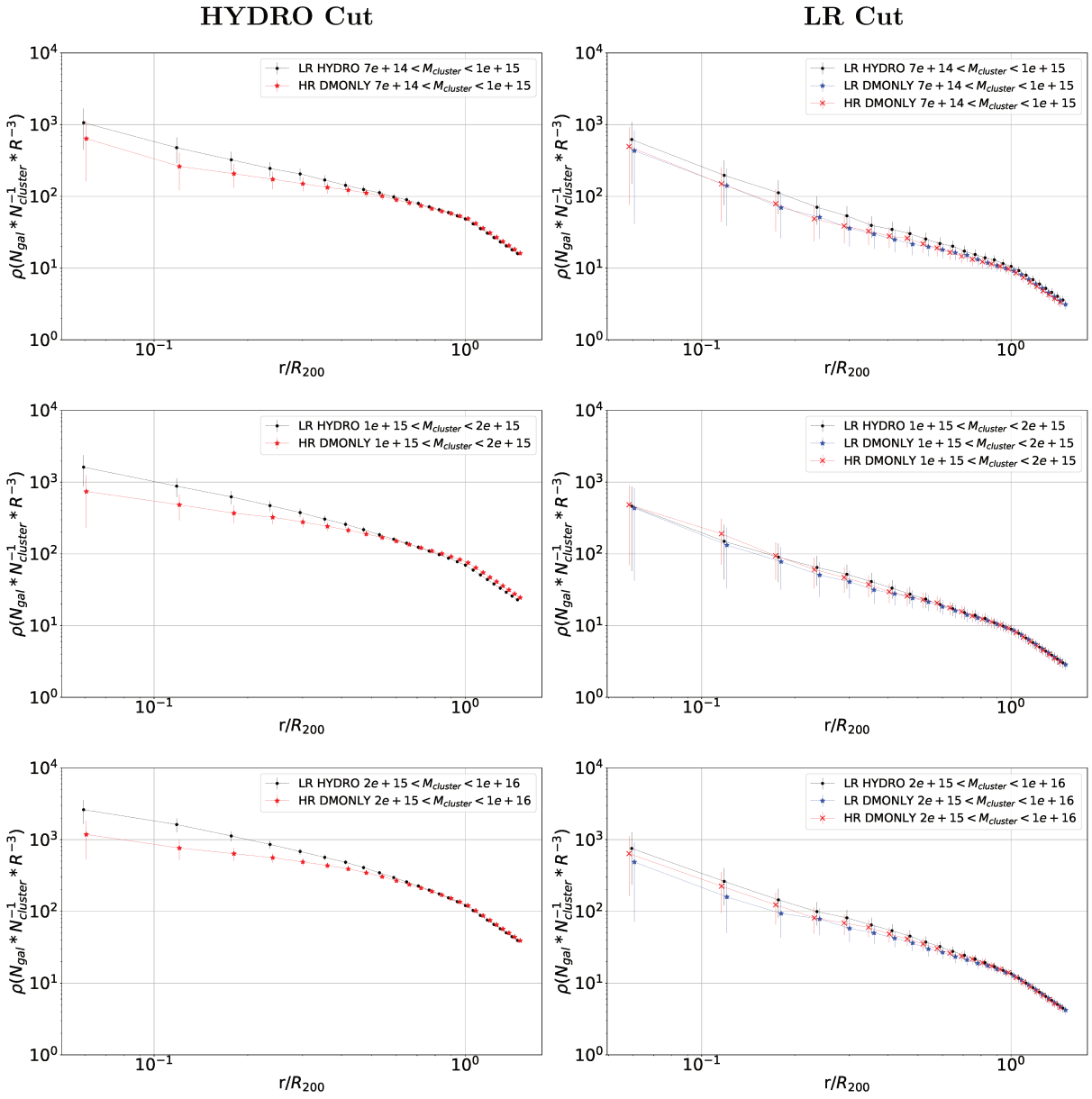


Figure 8.7 – 3D Cumulative galaxy density profiles. In left the resolution mass cut *HYDRO Cut* is applied to *LR HYDRO* and *HR DMONLY*. In the right the resolution mass cut *LR Cut* is applied to *LR HYDRO*, *LR DMONLY* and *HR DMONLY*. From top to bottom we show results for the three cluster mass bins (see Table 8.1): *Bin 1*, *Bin 2* and *Bin 3*, respectively. The data points are shifted radially for visualization purposes.

In all cases we observe a sharp decrease of the density after the $R_{200}$ of the cluster. This is because the number of galaxies is not changing, since in the 300th simulations there are no galaxies beyond $R_{200}$, but the volume is still increasing. So the density curve is falling as $1/r^3$,

which is linear in this figure representation. This region is not considered in the following.

### 8.3.2 Variation in Mass and Redshift

For the previous results we concentrated on redshift $z = 0$. However the redshift evolution of the galaxy density distribution is fundamental for understanding structure formation processes. Figure 8.8 shows the 3D cumulative galaxy density profiles when the *LR Cut* is applied, for the *LR HYDRO* and *LR DMONLY* simulations, in the left and right columns of the figure, respectively. From top to bottom we present results for the three cluster mass bins of Table 8.1: *Bin 1*, *Bin 2* and *Bin 3*, respectively. In each figure we have the redshift evolution following the color coding explained in Section 8.2.3. The data points are shifted in radius with respect to z = 0 for visualization purposes.

For the first cluster mass bin, we can see that even though we have applied the *LR Cut*, the *LR HYDRO* simulations present more galaxy density towards the center of the cluster where, as seen before, the small galaxies are located. This means that the *LR HYDRO* simulations keep smaller galaxies even when the resolution mass cut is applied.

In terms of the redshift evolution, we see the same behaviour for both simulations. At high redshift there are more galaxies towards the center, while in the outskirts the decrease is sharper. For low redshifts the change in the slope is smoother, having less structures in the center but more in the outskirts, and this feature is common for both simulations. This can be due to an increase of the fragmentation (bigger galaxies splitting into smaller ones) at high redshift. When the cluster mass bin increases, as we can see in the middle and bottom panels of the figure, the redshift differences are more clear, evidencing what we have said before. The significant differences in the profiles in the mass bin *Bin 3*, can be due to the low statistics at high redshift.

The main differences we can see for both simulations is the slope of the curves. For the *LR HYDRO* simulations the slope is higher, while for the *LR DMONLY* the curves tend to smooth towards the center. This lead us to conclude that the fragmentation takes place towards the center of the cluster where most of the collisions occur in the cluster formation in the hydrodynamical simulations, and due to the gas cooling effects, these collisions are reduced. Thus small galaxies are kept.

There is a clear difference between *LR HYDRO* and *LR DMONLY*, which are performed at the same resolution, so baryons affect the structure formation. Since the low resolution for the Three Hundred Project is lower than the resolution for *Euclid*, as we have seen in Chapter 5,

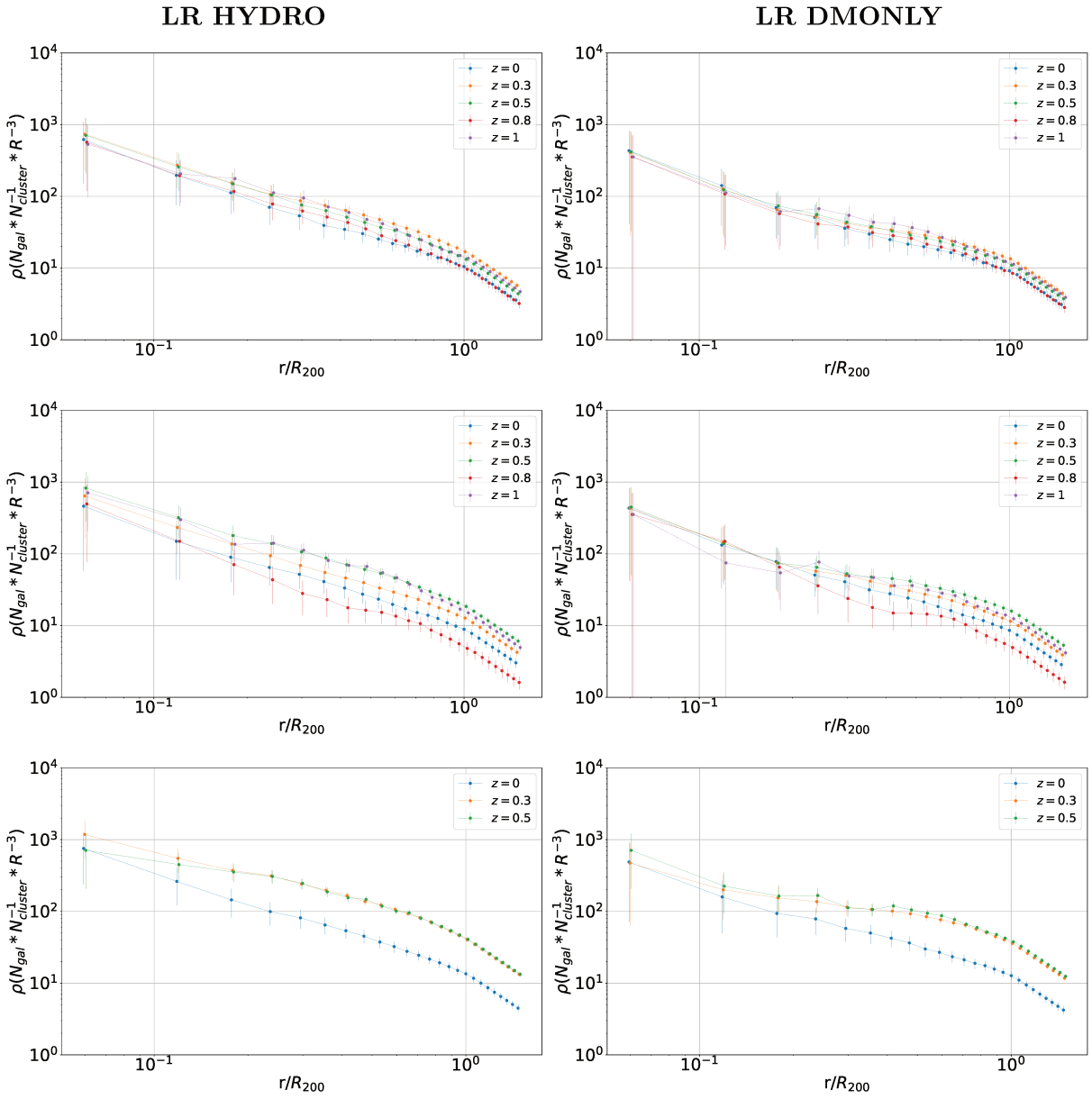


Figure 8.8 – 3D Cumulative galaxy density profiles for several redshifts between 0 and 1, when the resolution *LR Cut* is applied. The redshift color coding is as explained in Section 8.2.3. In left column we have the *LR HYDRO* simulations and in right the *LR DMONLY* ones. From top to bottom we present results for the three cluster mass bins (see Table 8.1): *Bin 1*, *Bin 2* and *Bin 3*, respectively. The data points are shifted radially with respect to z = 0 for visualization purposes.

we are more interested in checking the differences between *LR HYDRO* and *HR DMONLY*, and perform a more deep study of these profiles.

Figure 8.9 shows the 3D cumulative galaxy density profiles when the *HYDRO Cut* is applied, for the *LR HYDRO* and *HR DMONLY* simulations, in the left and right columns of the figure, respectively. From top to bottom we have the three cluster mass bins of Table 8.1. In each figure we have the redshift evolution following the color coding discussed in Section 8.2.3.

For the *Bin 1* cluster mass bin, we observe that even though we have applied the *HYDRO Cut*, the *LR HYDRO* simulations present more density towards the center where, as seen before, the small galaxies are located. This means that increasing the resolution in dark matter only simulations can not reproduce the baryons effect in the structure formation.

In terms of the redshift evolution, we see the same behaviour for both simulations. At higher redshift there are more galaxies towards the center, but decreasing rapidly in the outskirts. For low redshifts the change in the slope is smoother, having less structures in the center but more in the outskirts. This feature is common for both simulations. For the highest cluster mass bin, the behaviour is different for z = 0.5, and this can be due to the lack of statistics.

The main differences we can see for both simulations is the slope of the curves. For the *LR HYDRO* simulations the slope is higher, while for the *HR DMONLY* the curves tend to smooth towards the center. This lead us to same conclusion that for the low resolution case.

### 8.3.3 Modelling and Fitting

The goal is to compare the *LR HYDRO* and *HR DMONLY* simulations to see if the properties of the clusters can change. As there are resolution effects on the simulations, we need to account for them. We select to compare these two types of simulations to have a maximum of galaxies and be as realistic as possible. For this, we study, for both types of simulations, the 3D cumulative galaxy radial density profiles by fitting them with an analytical model and check the evolution of the parameters with mass and redshift, as performed in Chapter 5.

We use the *HYDRO Cut*, for being able to compare both simulations. The analytical model was chosen so that it can correctly fit the three types of simulations: *LR HYDRO*, *HR DMONLY* and *LR DMONLY*. It has been proven, by Navarro et al [106], that for numerical simulations with an accurate resolution, the NFW profile differs from the data, and that the Einasto profile is a better fit. Thus, we use the Einasto profile defined by [104, 105] as

$$\rho(r) = n_0 \exp\left(\frac{-2}{\alpha}\left[\left(\frac{r}{r_0}\right)^{\alpha} - 1\right]\right), \tag{8.2}$$

where $n_0$, $r_0$ and $\alpha$ are the free parameters describing the model. The behaviour of the model

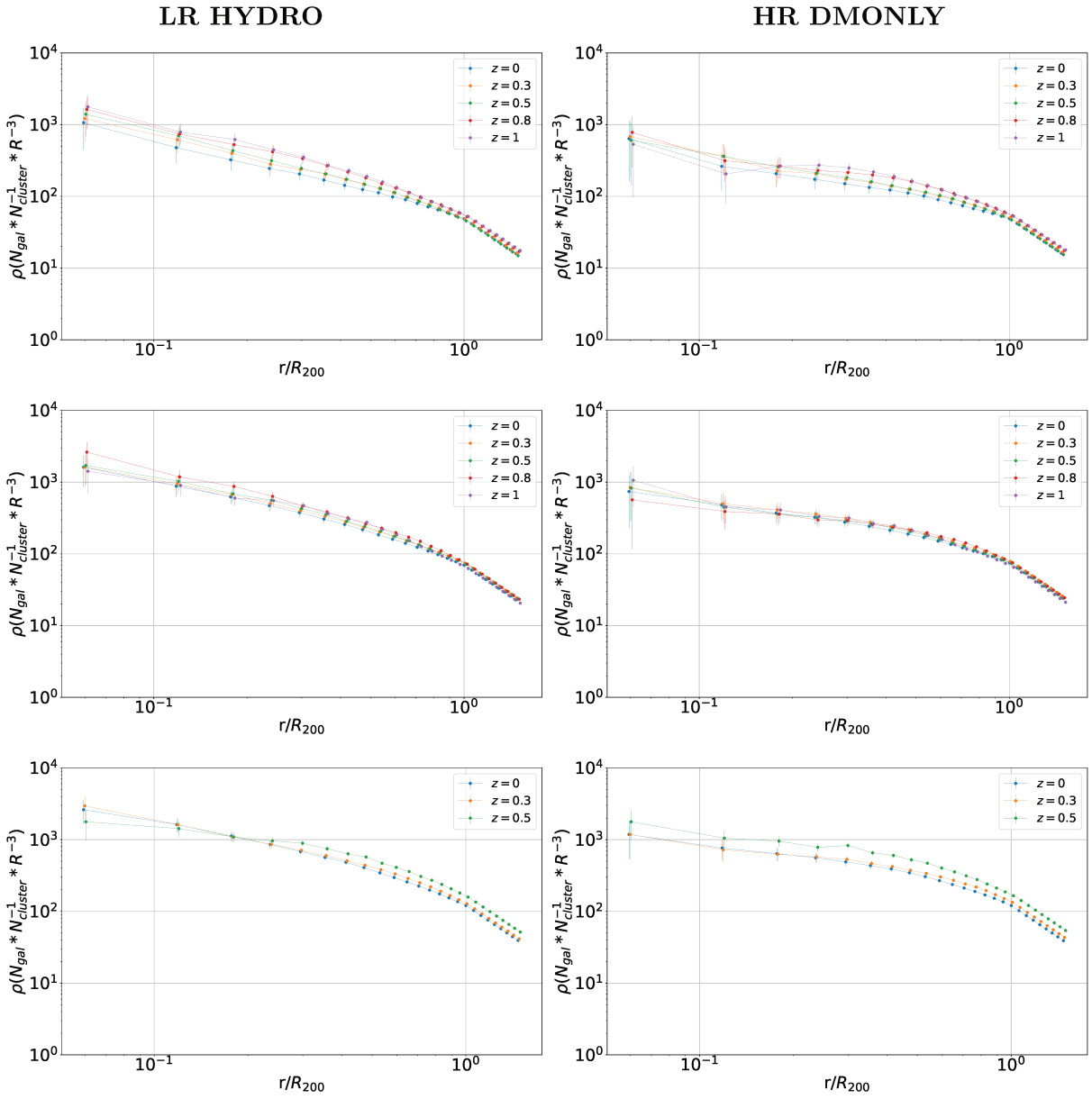


Figure 8.9 – 3D Cumulative galaxy density profiles for several redshifts between 0 and 1, when the resolution *HYDRO Cut* is applied. The redshift color coding is explained in Section 8.2.3. In left column we have the *LR HYDRO* simulations and in right the *HR DMONLY* ones. From top to bottom the three cluster mass bins (see Table 8.1): *Bin 1*, *Bin 2* and *Bin 3*, respectively. The data points for the different redshifts are shifted in radius with respect to z = 0 for visualization purposes.

with each parameter is presented in the following

- The parameter $n_0$ gives the normalization of the galaxy density. When the parameter

increases, so does the galaxy density.

- The parameter $\alpha$ affects the slope of the galaxy density distribution. When the parameter decreases, the slope increases.
- The parameter $r_0$ affects the curvature of the galaxy density profile. For a radius larger than $r_0$ the galaxy density profile drops rapidly.

We show in Figure 8.10 the Einasto model fit for the 3D cumulative radial galaxy density profiles for *LR HYDRO* and *HR DMONLY* in black and red (same color coding as before), respectively. The left column represents $z = 0$ and the right column, $z = 1$. From top to bottom two different cluster mass bins (Table 8.1: *Bin 1* and *Bin 2*, respectively. The black data points and its uncertainties are computed as explained in Section 8.3.1.

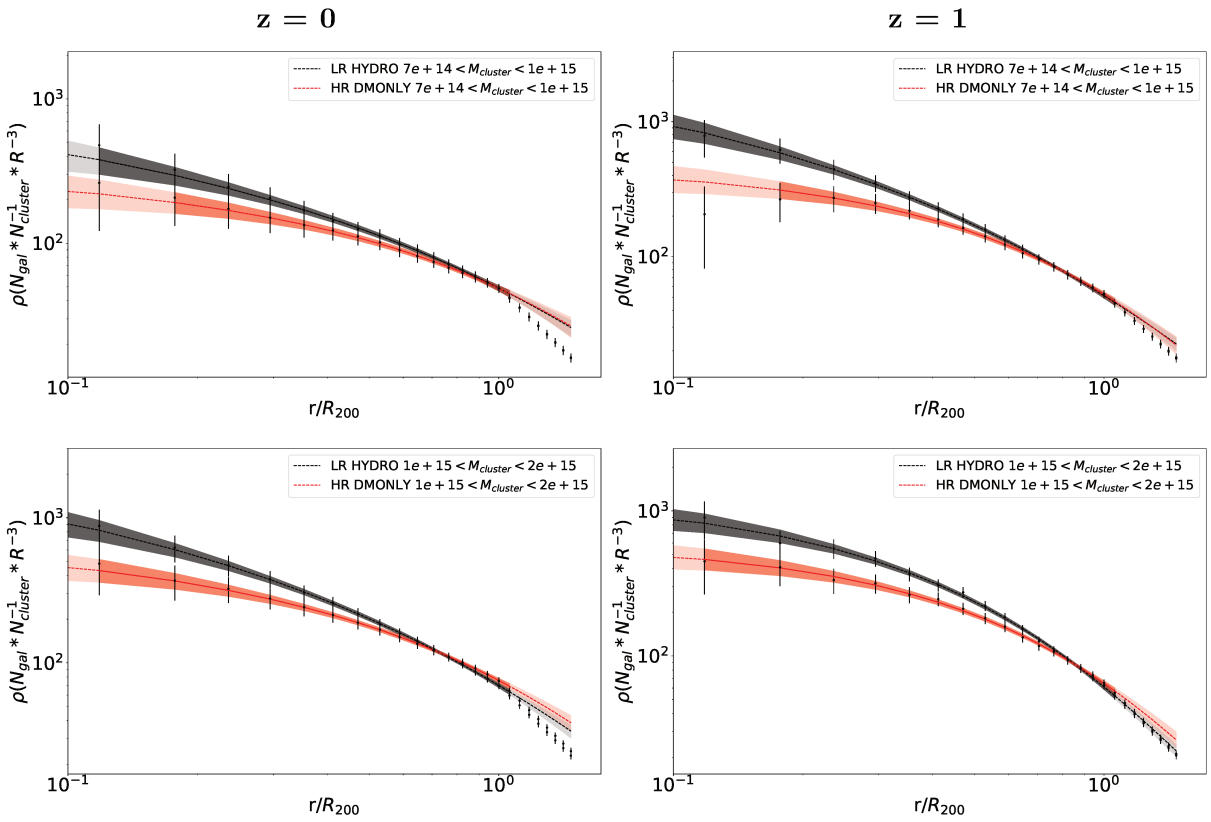


Figure 8.10 – Einasto fit for the 3D cumulative galaxy density radial profiles for the *HR DMONLY* and the *LR HYDRO* simulations when the *HYDRO Cut* is applied. The left column represents z = 0. The right column represents z = 1. From top to bottom we have *Bin 1* and *Bin 2* (see Table 8.1), respectively. The dotted lines and the shaded regions correspond to the best-fit value with its uncertainties based on the 16th, 50th and 84th percentiles over the posterior distribution. The way the fit and the data points with their uncertainties are computed is the same that in Chapter 5.

To fit the model we do not use the data beyond $R_{200}$, nor bins with less than three galaxies (see darkest regions in the Figure). The darkest regions are larger for *LR HYDRO* because we have more galaxies, which is in agreement with all the previous results. The model has been fitted by the same MCMC method explained in Chapter 5. The dotted lines and the shaded regions correspond to the best-fit value with its uncertainties based on the 16th, 50th and 84th percentiles over the posterior model distribution (median and 1 $\sigma$ spread). To perform the Einasto model fit in the cumulative distribution we have integrated the predicted number of galaxies on the spherical volume for each radius, r. As we can see in the figure, the Einasto profile is a very good fit to the data in all the cases.

We repeat the process for the three cluster mass bins and all the redshift bins we have. Later, we study the evolution of the three free parameters of the Einasto profile with mass and redshift. Due to the correlation found between the best-fit parameters, we will present the 2D posterior probability distributions found from the MCMC analysis.

Figure 8.11 shows the evolution of the parameters $\alpha$ and $r_0$ with the redshift for *LR HYDRO* in the left column, and *HR DMONLY* in the right one. The shaded colored regions represent the 68 % C.L. We show results for the cluster mass bins *Bin 1*, *Bin 2* and *Bin 3* from top to bottom, respectively. We can see in the figure that the values of $r_0$ and $\alpha$ decrease with increasing redshift. Furthermore, the uncertainties decrease. This can be due to the fact that there are less clusters at high redshift, so the scatter is lower. However, there is a clear tendency for the evolution of the parameters. With respect to the simulation type, for *HR DMONLY* both parameters show a wider value range, and in general, the $\alpha$ value is higher. With respect to $r_0$, we observe that the value is consistent for both types of simulations, while $\alpha$ keeps being greater at high mass for *HR DMONLY*. These results are in agreement with the profiles shown in Figure 8.9. At higher redshift the slope of the galaxy density profiles is higher. Following the behaviour of the Einasto model with its parameters (explained before), if the $\alpha$ and $r_0$ parameters decrease with redshift, means a higher slope towards the center that decrease rapidly towards the outskirts of the cluster, equivalently a more concentrated cluster. In addition, in Figure 8.9, the density is higher for *LR HYDRO* than for *HR DMONLY*. For this reason the value of $\alpha$ is higher for the latter, because its slope is smoother towards the center.

Figure 8.12 is equivalent to Figure 8.11 but for the parameters $r_0$ and $n_0$. We observe that these parameters are highly correlated and in a similar way like for the $\alpha$ and $r_0$ and for both types of simulations. For *LR HYDRO* we have already explained in previous the section that baryon cooling allows to preserve structures that dark-matter-only simulations do not. In terms of the redshift and mass, the $r_0$ parameter decreases, and $n_0$ increases. The fact that $n_0$ increases with redshift can be due to, as explained in Section 8.3.2, that at high redshift there is more

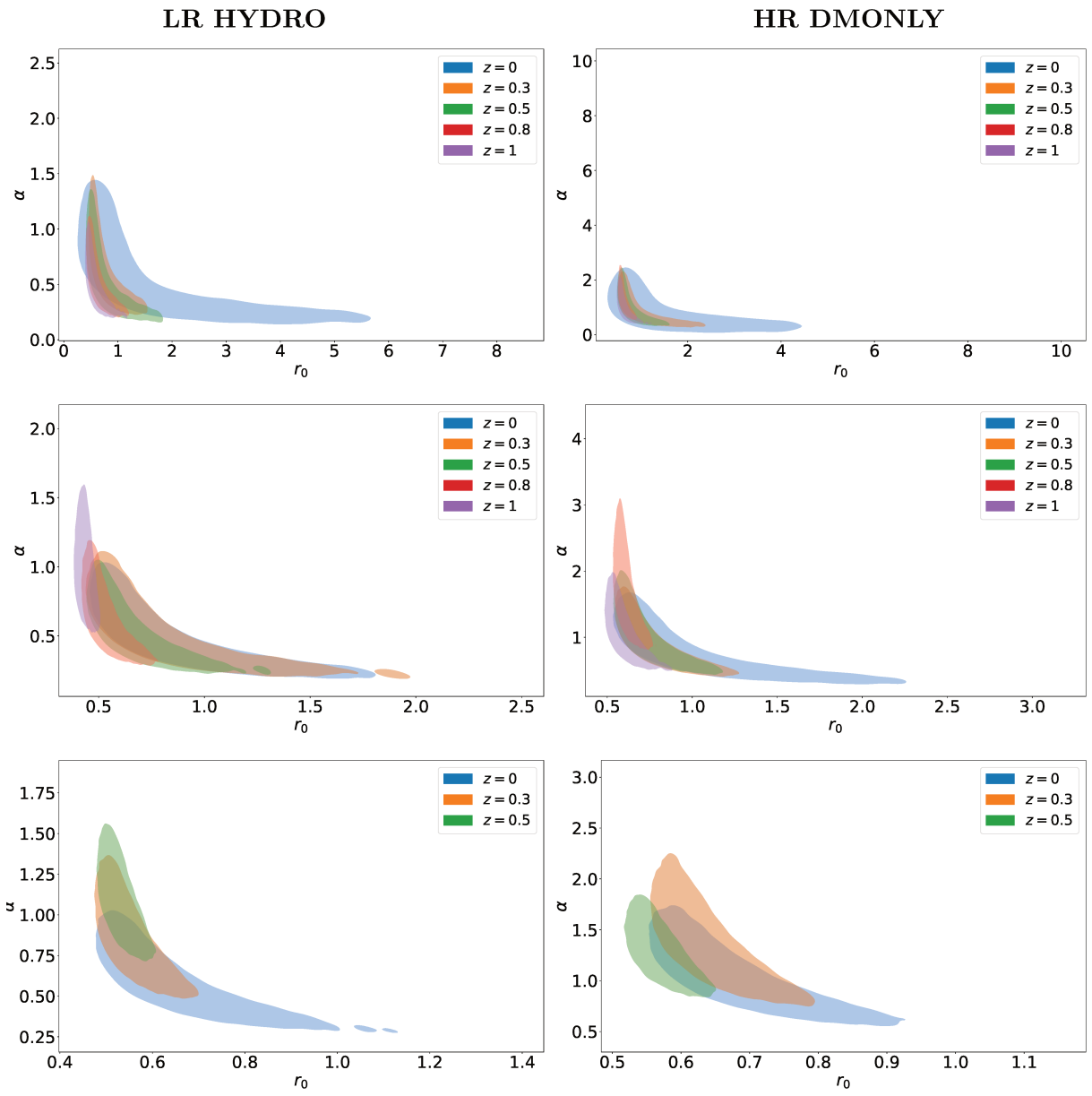


Figure 8.11 – 2D probability distribution of the $\alpha$ and $r_0$ parameters of the Einasto model with redshift. In the left column we have the *LR HYDRO* simulations. In the right column we have the *HR DMONLY* simulations. From top to bottom we have three different mass bins, following Table 8.1, *Bin 1*, *Bin 2* and *Bin 3*, respectively. The shaded regions are the 68 % C.L.

fragmentation, then a larger number of galaxies is giving a higher density value. These results are in agreement with the profiles shown in Figure 8.9. At higher redshift galaxy density is larger but it decreases rapidly towards the outskirts. Following the behaviour of the Einasto model with its parameters (explained before), if the $n_0$ increases, so does the galaxy density. If $r_0$ decreases, the galaxy density drops rapidly for radius larger than $r_0$. Figure 8.9, the density

is higher for *LR HYDRO* than for *HR DMONLY*. For this reason the value of $n_0$ is larger for the latter.

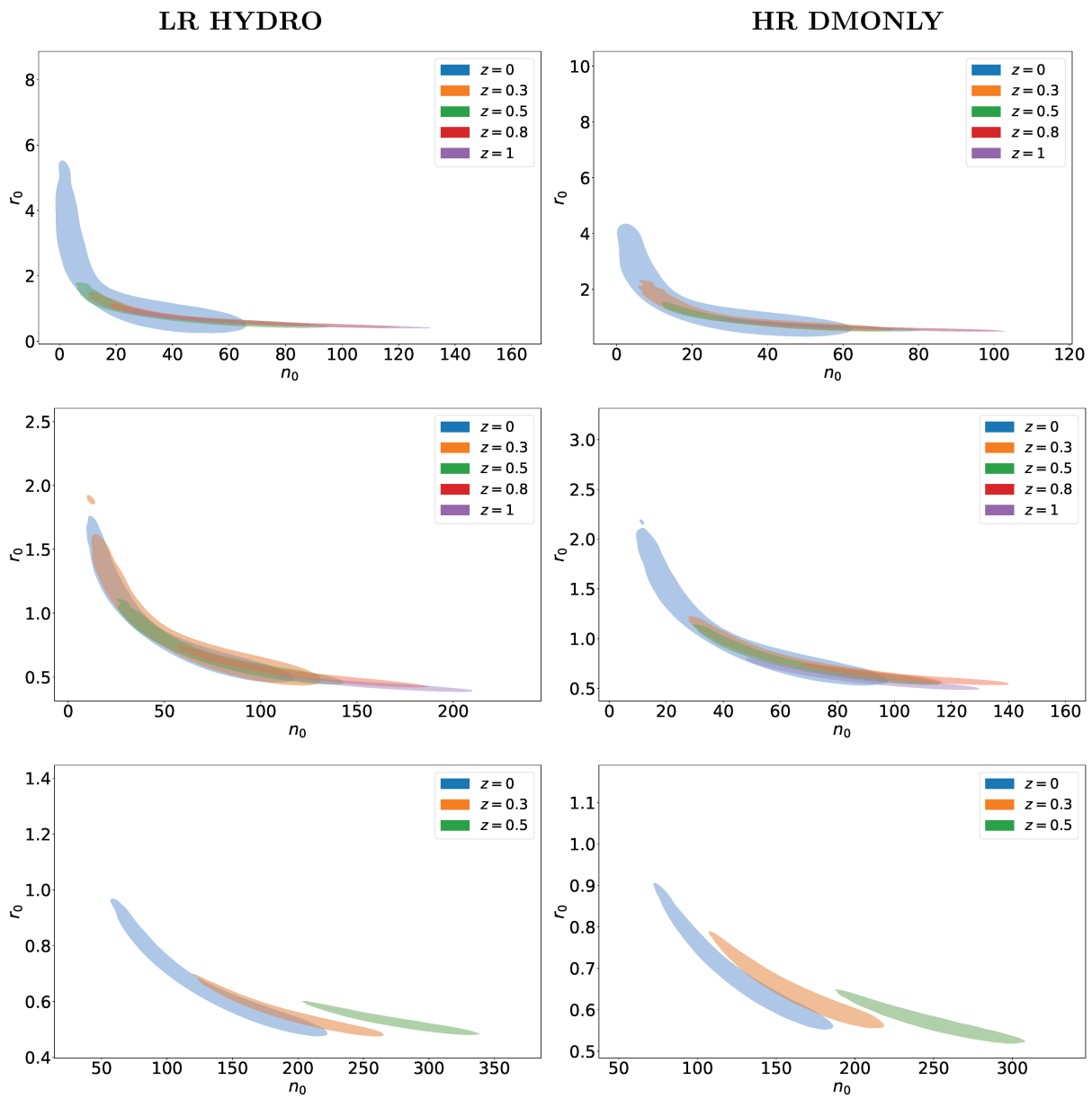


Figure 8.12 – 2D probability distribution of the $r_0$ and $n_0$ parameters of the Einasto model with redshift. In the left column we have the *LR HYDRO* simulations. In the right column we have the *HR DMONLY* simulations. From top to bottom we have three different mass bins, following Table 8.1, *Bin 1*, *Bin 2* and *Bin 3*, respectively. The shaded regions are the 68 % C.L.

Finally, Figure 8.13 shows the 2D posterior probability distribution for the parameters $\alpha$ and $n_0$. We observe that both parameters increase with mass and redshift, having a different

degeneracy with respect to the previous cases. The main behaviour we observe is that even with the same resolution mass cut, both simulations shows different parameter values. The $\alpha$ parameter, the curvature, is higher for the *HR DMONLY* but the $n_0$ is lower. The fact that $\alpha$ is higher and $n_0$ is lower for the *HR DMONLY* simulations, tell us that the galaxy density is lower, as expected. In addition, $\alpha$ decreases with redshift, while $n_0$ increases, for both types of simulations. So these results are also in agreement with the previous ones in Figure 8.9. We can conclude that the Einasto model performs a good fit of the data because the redshift and mass evolution of its parameters reproduce the data behaviour.

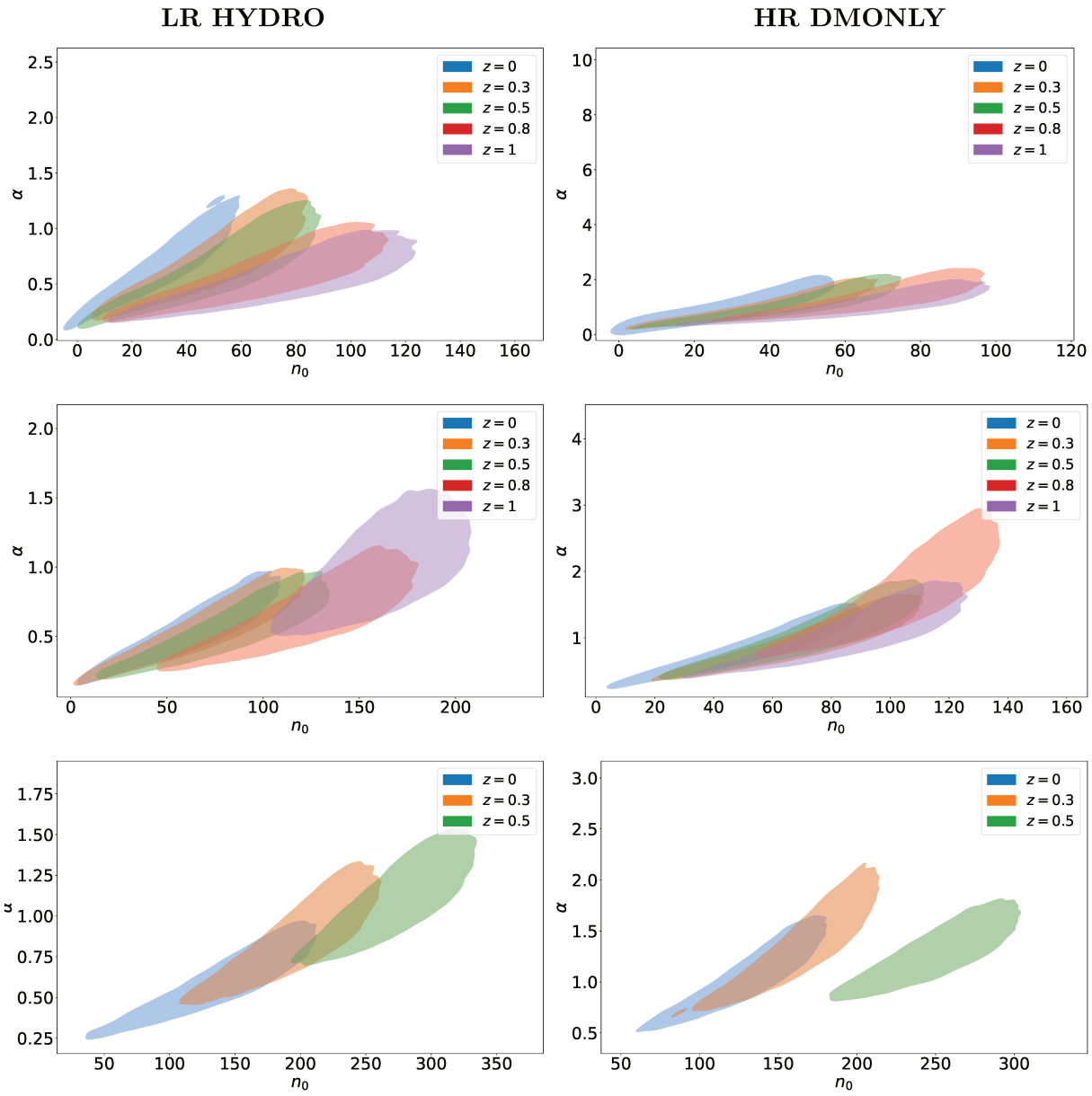


Figure 8.13 – 2D probability distribution of the $\alpha$ and $n_0$ parameters of the Einasto model with redshift. In the left column we have the *LR HYDRO* simulations. In the right column we have the *HR DMONLY* simulations. From top to bottom we have three different mass bins, following Table 8.1, *Bin 1*, *Bin 2* and *Bin 3*, respectively. The shaded regions are the 68 % C.L.

## 8.4 A possible application of the Three Hundred results for *Euclid*

In previous sections we have discussed the effects of resolution on the galaxy density distribution, and in the LF. We have observed how the inclusion of baryonic physics in numerical simulations affect significantly both observational properties. Therefore, the main interest of this work lies in its possible application to construct a more realistic cluster catalogue for the estimation of the selection function for *Euclid*. The major problem with the catalogues we have used in this chapter is the lack of statistics with respect to the amount of cluster that *Euclid* will detect. For example, for the 68 regions we have studied, there are no central clusters for z $> 0.5$ at large cluster masses.

To improve the statistics with the current catalogue we could

- Use the 324 regions for the *LR HYDRO* simulations instead of 68 regions. The latter were used to establish a resolution mass cut. However, this resolution mass cut may be applied to the 324 regions. This will enlarge our dataset and, therefore, our statistics.
- We could use more than the central cluster in the regions. This allows us to have more clusters at low mass and at high redshift.

For the LF the *HR HYDRO* simulations are needed. In this thesis we have demonstrated that even just for a single cluster the LF can be well determined. The Three Hundred collaboration is currently running those simulations and we expect to have a larger sample soon.

For now one could imagine to adopt a hybrid methodology by including changes in the galaxy density distribution to approach the properties observed in the Three Hundred simulations while keeping the current *Euclid* LF until more *HR HYDRO* simulations are available.

## 8.5 Conclusions and Perspectives

In this chapter we have studied the optical cluster properties of the 300th Cluster catalogue between hydrodynamical and N-body simulations, accounting for resolution effects.

At the beginning, The main goal of this analysis was to repeat the analysis done for EUCLID in Chapter 5. However, we computed the luminosity function (LF) for hydrodynamical simulations with lower resolution than the *Euclid* N-body simulations, and we observed a lack of galaxies in the faint part of the LF. When computing the LF for one high resolution hydrodynamical cluster, this problem was solved. For this reason we decided to perform an analysis of resolution effects in cluster properties for hydrodynamical and N-body simulations.

We computed the galaxy mass function and the galaxy density radial distribution. We fitted these properties with analytical models and we studied the distribution with mass and redshift of the free parameters of each model. Observing the results we conclude that the choice of the bins is not right. We chose the bins at redshift zero, for ensuring that cluster properties within a bin are similar while having a sufficient number of clusters per bin. However, at high redshift the number of clusters drop. For the future, it would be better to arrange the three bins in two bins, to ensure larger statistics but preserving the cluster physical properties.

When comparing N-body simulations for two different resolutions, we realized that increasing the resolution lead to fragmentation of galaxies. This means that, massive galaxies in the low resolution simulations are divided into smaller ones in the high resolution simulations. This leads to larger values of the galaxy mass function and the galaxy density. For hydrodynamical simulations, we conclude that they keep small galaxies alive possibly due to cooling processes of gas in the simulations. These galaxies are generally located towards the center of the cluster. When increasing the resolution of N-body simulations, this feature can not be replicated. At higher redshift, this behaviour is more significant, because we find more fragmentation. Thus the galaxy density and galaxy mass function is larger at higher redshift, and mass.

The fact that baryonic physics affect the structure formation processes may affect the cluster finders performances. As seen in Chapter 5, the luminosity function does not change with the cluster mass, but with its distance. Therefore, having more fragmentation, and keeping smaller galaxies, instead of overmerging them, may affect cluster finders such as PZWAV, which looks for overdensities when identifying a cluster.

# CONCLUSIONS AND PERSPECTIVES

My thesis was carried out within two main collaborations: *Euclid* [41] and The Three Hundred [43], with two main topics

1. Characterization of the readout noise of the *Euclid* infrared detectors in the NISP instrument.
2. Study observational optical/infrared properties of galaxy clusters for the estimation of the selection function.

Infrared instruments, and in particular the NISP, acquire data using the MACC readout mode, which consists of a series of non-destructive exposures averaged into groups that form a ramp. The input flux in the detectors can then be obtained from the slope of the ramp using maximum likehood estimators, which generally assume white readout noise. We have extended these estimators to the case of correlated readout noise. Analytical expressions for the group and group difference covariance matrices are presented for the case of $(1/f)^{\alpha}$-like correlated readout noise. These have been validated via Monte Carlo simulations.

Furthermore, we have studied the readout noise associated to NISP detectors taking advantage of long exposure (few hours) performed during laboratory dark tests at the CPPM cryogenic facilities. We have found that the NISP readout noise is mildly correlated and can be well characterized by a $(1/f)^{\alpha}$-like model. From this we conclude that the readout noise of the NISP detectors has non-negligible correlation within the typical in-flight NISP exposure time (574 seconds).

Finally, we have performed Monte Carlo simulations of the in-flight expected NISP detector signal and noise, including a realistic background signal and correlated readout noise as measured on the ground calibration tests. From these simulations we have been able to estimate the expected bias in the on-board flux estimates during in-flight operations for which white readout noise is assumed. We find that for the spectroscopic mode of the NISP instrument, NISP-S, low background the flux bias can be up to four times larger than when accounting for the correlation in the readout noise. Nevertheless, this bias is negligible for typical sky background signals. Therefore, we expect no significant bias in the on-board fluxes measured by EUCLID. On the other hand, for the photometric mode, NISP-P, the white noise approximation is the one that should be taking into account since the exposure time of the NISP-P mode corresponds to a frequency domain where the power spectrum is dominated by a white noise. This work is published

in Publications of the Astronomical Society of the Pacific (PASP) as Jiménez Muñoz et al. [201].

The second part of the thesis includes a work done in the context of cosmology with galaxy clusters. The final goal was to determine the *Euclid* selection function through the cluster injection method [202]. To carry out this project, first, we have studied the two most important observational cluster properties using the 300 deg$^2$ *Euclid* Mock catalogue [134]: the galaxy density radial profile and the luminosity function. The first of the two properties was recovered from the mock catalogue, creating bins in clusters' mass and redshift, to later be fitted by a truncated analytical Navarro-Frenk-White (NFW) [101] model following previous studies. Overall, we find that this model does not fit properly the data. However, we have demonstrated during this thesis that for dark-matter-only simulations, as it is the case for the *Euclid* Mock catalogue, and for hydrodynamical simulations, the galaxy density distribution is well-fitted by an Einasto profile [104, 105] at $1\sigma$. This Einasto model should then be used for the analysis of the real *Euclid* data.

The luminosity function (LF) is computed using the *Euclid* Mock catalogue, for the same bins in mass and redshift, to later be fitted to a Schechter model [93]. Generally, this model does not fit the brightest region of the LF while it is a good fit to the fainter region. To perform a good fit to the brightest region it would suffice to establish a maximum threshold in magnitude for the galaxies, but this will need to be completed with another model to consider the faintest part. For our purposes we did not consider this cut because we wanted to reproduce the full LF. For future works we may consider to chose a different analytical model, for example a double Schechter function [203], to improve the fit. The free parameters for both models were recorded for each bin in mass and redshift. Overall, the binning choice could be improved by choosing narrower regions, specially in redshift. In summary, the luminosity function is not well recovered by the Schechter model while the galaxy density distribution, as a first approximation, may be described by a NFW profile. However, as a perspective we consider to explore other analytical models that recover the *Euclid* cluster's observational properties in a more precise way.

The next step was to construct a synthetic cluster catalogue for cluster injection. To avoid computational cost problems and as a first attempt, we decided to construct a portion of 36 deg$^2$ on the sky within the limits of the 300 deg$^2$ *Euclid* Mock catalogue. We generated the expected number of clusters for each bin in mass and redshift randomizing the mass and redshift within the bin limits. The cluster galaxy members density distributions of the synthetic clusters were constructed from the NFW profiles using the parameters stored from the fits discussed above. The luminosity function was chosen to be the same of the *Euclid* catalogue because, as discussed, the Schechter model does not describe fully the data. In practice we just draw the magnitude of the cluster galaxy members from the binned LF for each bin in mass and redshift. The cluster

member galaxies' redshifts follow a Gaussian distribution with the cluster redshift as the mean, and as standard deviation, the instrumental *Euclid* photometric uncertainies as given by the *Euclid* Red Book [41].

The number of member galaxies for clusters in the same bin in mass and redshift was the same. As a whole, the synthetic clusters preserve the main properties of the ones in the *Euclid* Mock catalogue. However, they could be improved by adding several properties we did not consider, such as: velocity dispersion, ellipticity, dispersion for the number of galaxies or photometric redshift dispersion other than the *Euclid* instrumental uncertainties. To have a test catalogue of galaxies to compare to the *Euclid* one we also simulated field galaxies from the Schechter model luminosity function, and a standard redshift distribution from Chang et al. [187].

The synthetic clusters were then injected into the *Euclid* Mock catalogue. We have performed two analysis: (i) The synthetic clusters are directly injected into the *Euclid* Mock catalogue without removing the detected clusters in the 36 $\deg^2$ region, (ii) The synthetic cluster catalogue is injected into the *Euclid* Mock catalogue for which all the galaxies have been randomly shuffled in space and redshift, so that the original clusters are "removed". The main issue with respect to the first method is the overdensity of galaxy clusters that can overlap. This can hinder the detection of galaxy clusters by the cluster finder. As for the second method, the fact that we randomize the galaxy positions breaks the spatial correlation between structures which may affect cluster detection with respect to the original *Euclid* Mock catalogue. In addition, there is an overdensity of field galaxies in the 36 $\deg^2$ region because we did not remove Mock *Euclid* galaxy cluster members. Overall, both methods could be improved by explicitly removing the detected cluster galaxy members. However, the first method allows us to check if the different nature of galaxy clusters affect the cluster finders, while the second method enables us to verify if the environment plays a key role in the detection of galaxy clusters.

To compute the selection function is essential to characterize the performance of a cluster finder through the completeness and purity. The PZWAV cluster finder [134] has been ran in the four catalogues discussed above: 1) *Euclid* Mock catalogue, 2) synthetic cluster and field galaxies catalogue, 3) synthetic clusters injected in the *Euclid* Mock catalogue and 4) synthetic clusters injected in the shuffled *Euclid* Mock catalogue. To calculate the matchings between the output of PZWAV and the simulated catalogues we used a geometrical matching procedure by defining a volume in redshift and cluster distance [134]. We have checked that the distance in redshift is mainly affected by the environment but not by the properties of the cluster. On the other hand, the volume distance transversal to the line of sight depends on the size of the clusters, thus on their physical properties. Overall, the completeness and purity from PZWAV is greater for the synthetic cluster and field galaxies catalogue. The worst performance occurs

for the injection catalogue including the *Euclid* Mock clusters. This tells us that the difference in the field galaxy spatial structure and the presence of other clusters affects significantly the detection performance. To check the effect from the environment we compared the synthetic catalogue with the injection after randomizing the galaxy position and redshift. We observe that PZWAV finds more true clusters that in the case of the other injection catalogue. Thus, the cluster overdensities are clearly affecting the detections. In addition, with respect to the full synthetic catalogue the PZWAV performance worsen. We conclude that, the environment plays a key role. Comparing the results between the *Euclid* Mock catalogue and the synthetic catalogue, we observe that simulated clusters with basic properties are more likely to be detected than the ones in the *Euclid* Mock catalogue.

In terms of number of galaxies, the synthetic clusters are observed with less richness than the *Euclid* ones. At the time of writing this thesis it is not clear why this is happening, and why the richness of the detected synthetic cluster depends strongly on the environmental properties. To go further in the analysis we would like to include more cluster properties and modify the ones that we have simulated with models that fit better the data. For the injection method, it would be also interesting to remove the galaxy clusters in the *Euclid* Mock catalogue and inject the synthetic ones at the same positions, to keep the spatial correlation between structures, and avoid "fake" overdensities. In addition, we have illustrated in this thesis the complexity of constructing a selection function either using a mock catalogue or an injection catalogue methodology. We have shown that the properties of the field galaxies in terms of the spatial correlation plays a key role and need to be reproduce accurately in both methodologies. Further work is needed to better define how to go from the properties of the real survey detected clusters to a simulation of synthetic clusters.

The final project presented in this thesis was in the framework of the Three Hundred Collaboration [43], for which we started a collaboration with M. De Petris, G. Yepes, W. Cui and A. Ferragamo. The Three Hundred is a sample of 324 cluster regions that have been re-simulated with N-body simulations and with hydrodynamical simulations [54], for several mass resolutions. The main goal of this collaboration was to construct a synthetic catalogue, as discussed above, with more realistic cluster properties from the hydrodynamical simulations. For the latter, we checked that at high resolution, an Einasto profile and a Schechter model perform a good fit to the hydrodynamical simulations. However, the lack of statistics for high resolution for hydrodynamical simulations (only one region was available) did not enable us to generate a full synthetic cluster catalogue. We have also studied the low resolution hydrodynamical simulations. We have found that in this case it is not possible to derive the LF, because of lack of resolution we are limited to an apparent magnitude about 21 in the H band. As a consequence,

we decided to concentrate only on the galaxy distribution and study the resolution effects on the galaxy density profile, which can be computed for both hydrodynamical and dark-matter-only simulations. We observed that low mass galaxies are generally located towards the center of the clusters. Furthermore, hydrodynamical simulations tend to preserve small structures, that can not be found in dark-matter-only simulations, even when the resolution of the latter is improved significantly. Thus, the effect of baryonic physics in the structure formation plays a key role in the observational properties of galaxy clusters, which may affect the cluster finders' performance.

In general, we could improve the choice of binning in mass and redshift, because we have a lack of statistics at high mass and redshift. Moreover, we could have considered other clusters apart from the central one for the 324 regions, to explore a wider range in cluster mass and redshift and improve the statistics. Furthermore, the significant computational cost for high resolution hydrodynamical simulations avoids the possibility of constructing a luminosity function. However, we could construct a synthetic catalogue by mixing realistic properties, as the galaxy density profile from the Three Hundred Project, with the luminosity functions coming from the *Euclid* Mock catalogue, or extrapolate the low resolution LF to fainter magnitudes.

The final objective to be reached would be to compute the selection function in the most realistic way. This will allow us to do cosmology with galaxy clusters in the context of *Euclid*, calculating the cluster abundance and constraining mainly the cosmological parameters $\sigma_8$ and $\Omega_m$ [137]. For this thesis we have performed a step forward in that direction but work will be need to fully exploit the future *Euclid* Cluster catalogue.

Appendix A

# Detailed computation of the group readout noise covariance matrix

We detail here the computation of the group noise covariance matrix for correlated readout noise. We concentrate in the correlated readout noise terms. Other terms can be found in [168].

The group noise covariance matrix is given by:

$$C_{kk} = (k-1)\,\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)(2m+1)}{6m} + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\delta\rho_i^{(k)}\delta\rho_j^{(k)},$$

for the diagonal term, and:

$$C_{kl} = <\delta G_k \delta G_l> = (k-1)\,\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)}{2} + \left\langle \left[\frac{1}{m}\sum_{i=1}^{m}\delta\rho_i^{(k)}\right] * \left[\frac{1}{m}\sum_{j=1}^{m}\delta\rho_j^{(l)}\right]\right\rangle,$$

for the off-diagonal elements. Then we compute the expressions for the readout noise:

- $$\left\langle \left[\frac{1}{m}\sum_{i=1}^{m}\delta\rho_i^{(k)}\right] * \left[\frac{1}{m}\sum_{j=1}^{m}\delta\rho_j^{(l)}\right]\right\rangle = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} < \rho_i^{(k)}\rho_j^{(l)} >= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C_{[l*(m+d)+j,k*(m+d)+i]} =$$

$$\frac{1}{m^2}\sum_{i=1}^{m} C_{[l*(m+d)+i,k*(m+d)+i]} + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C_{[l*(m+d)+j,k*(m+d)+i]}$$

$$= \frac{1}{m^2}\sum_{i=1}^{m} C[|(l-k)*(m+d)+(i-i)|*t_{frame}]+$$

$$\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C[|(l-k)*(m+d)+(j-i)|*t_{frame}] =$$

$$\frac{1}{m^2} * mC[(l-k)*(m+d)*t_{frame}]+$$

$$\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C[|(l-k)*(m+d)+(j-i)|*t_{frame}] =$$

$$\frac{1}{m^2}\left[mC[(l-k)*(m+d)*t_{frame}] + \sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C\left[|(l-k)*(m+d)*+(j-i)|*t_{frame}\right]\right]$$

• Diagonal terms, so $l = k$ , then:

$$\frac{1}{m^2}\left[mC[(l-k)*(m+d)*t_{frame}] + \sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C\left[|(l-k)*(m+d)*t_{frame}+(j-i)|*t_{frame}\right]\right] =$$

$$\frac{1}{m^2}\left[mC[0] + \sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C\left[|(j-i)|*t_{frame}\right]\right]$$

$$= \frac{1}{m^2}\left[mC(0) + 2\sum_{i=1}^{m-1}(m-i)C(i*t_{frame})\right]$$

$$C_{kk} = (k-1)\,\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)(2m+1)}{6m} + \frac{1}{m^2}\left[mC(0) + 2\sum_{i=1}^{m-1}(m-i)C(i*t_{frame})\right]$$

• Off-Diagonal Matrix, so $l \neq k$ , then:

$$\left\langle \left[\frac{1}{m}\sum_{i=1}^{m}\delta\rho_i^{(k)}\right] * \left[\frac{1}{m}\sum_{j=1}^{m}\delta\rho_j^{(l)}\right]\right\rangle = \\ +\frac{1}{m^2}\left[mC[(l-k)*(m+d)*t_{frame}]\right. \\ \left.+\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C\left[|(l-k)*(m+d)*t_{frame}+(j-i)|*t_{frame}\right]\right]$$

$$C_{kl} =< \delta G_k \delta G_l >= (k-1)\,\mathcal{D} + (k-1)(m-1)f + f\frac{(m+1)}{2} \\ +\frac{1}{m^2}\left[mC[(l-k)*(m+d)*t_{frame}]\right. \\ \left.+\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m} C\left[|(l-k)*(m+d)*t_{frame}+(j-i)|*t_{frame}\right]\right]$$

## Group differences

### Diagonal Terms

$$D_{kk} = d + f\frac{(m-1)(2m-1)}{3m} \\ +\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle\delta\rho_i^{(k+1)}\delta\rho_j^{(k+1)}\right\rangle + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle\delta\rho_i^{(k)}\delta\rho_j^{(k)}\right\rangle \\ -\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle\delta\rho_i^{(k+1)}\delta\rho_j^{(k)}\right\rangle - \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle\delta\rho_i^{(k)}\delta\rho_j^{(k+1)}\right\rangle$$

We compute now each term of the correlated readout noise contribution:

• $$\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle\delta\rho_i^{(k+1)}\delta\rho_j^{(k+1)}\right\rangle$$

$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle\delta\rho_i^{(k)}\delta\rho_j^{(k)}\right\rangle = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|j-i|*t_{frame}]$$

$$\bullet \quad \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k+1)}\delta\rho_j^{(k)}\right\rangle$$

$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(k-(k+1))*(m+d)+(j-i)|*t_{frame}]$$

$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(j-i)-(m+d)|*t_{frame}]$$

$$\bullet \quad \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k)}\delta\rho_j^{(k+1)}\right\rangle$$

$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|((k+1)-k)*(m+d)+(j-i)|*t_{frame}]$$

$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(j-i)+(m+d)|*t_{frame}]$$

Finally, the group difference diagonal covariance matrix is:

$$D_{kk} = \mathcal{D} + f\frac{(m-1)(2m-1)}{3m} +$$
$$2\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|j-i|*t_{frame}]$$
$$-\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(j-i)+(m+d)|*t_{frame}]$$
$$-\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(j-i)-(m+d)|*t_{frame}]$$

**Off-diagonal terms**

$$D_{kl} = \frac{f}{6m}(m^2-1)\delta_{(k+1)l}$$
$$+\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k+1)}\rho_j^{(l+1)}\right\rangle + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k)}\rho_j^{(l)}\right\rangle$$
$$-\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k+1)}\rho_j^{(l)}\right\rangle - \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k)}\rho_j^{(l+1)}\right\rangle$$

We compute now each term of the correlated readout noise contribution:

$$\bullet \quad \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k+1)}\rho_j^{(l+1)}\right\rangle = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k)}\rho_j^{(l)}\right\rangle$$
$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-k)*(m+d)+(j-i)|*t_{frame}]$$

$$\bullet \quad \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k+1)}\rho_j^{(l)}\right\rangle$$
$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-(k+1))*(m+d)+(j-i)|*t_{frame}]$$
$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-k-1)*(m+d)+(j-i)|*t_{frame}]$$

$$\bullet \quad \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \delta\rho_i^{(k)}\rho_j^{(l+1)}\right\rangle$$
$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|((l+1)-k)*(m+d)+(j-i)|*t_{frame}]$$
$$= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-k+1)*(m+d)+(j-i)|*t_{frame}]$$

Finally, the group difference off-diagonal covariance matrix is:

$$D_{kl} = \frac{f}{6m}(m^2-1)\delta_{(k+1)l}$$
$$+2\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-k)*(m+d)+(j-i)|*t_{frame}]$$
$$-\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-k-1)*(m+d)+(j-i)|*t_{frame}]$$
$$-\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} C[|(l-k+1)*(m+d)+(j-i)|*t_{frame}]$$

# BIBLIOGRAPHY


[1] A Einstein. The foundation of the general theory of relativity.[first published in 1916 as die grundlage der allgemeinen relativitätstheorie, annalen der physik, 354, 769-822]. *W. Perrett & GB Jeffery (Trans.), The Principle of Relativity: A Collection of Original Papers on the Special and General Theory of Relativity*, pages 109–164, 1916.

[2] Alexander Friedman. Über die krümmung des raumes. *Zeitschrift für Physik*, 10(1):377–386, 1922.

[3] Alexander Friedmann. Über die möglichkeit einer welt mit konstanter negativer krümmung des raumes. *Zeitschrift für Physik*, 21(1):326–332, 1924.

[4] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles,

A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. , 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

[5] Georges Lemaître. Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. In *Annales de la Société scientifique de Bruxelles*, volume 47, pages 49–59, 1927.

[6] Georges Lemaître. The beginning of the world from the point of view of quantum theory. *Nature*, 127(3210):706–706, 1931.

[7] Georges Lemaître. L'expansion de l'espace. *Publications du Laboratoire d'Astronomie et de Geodesie de l'Universite de Louvain, vol. 8, pp. 101-120*, 8:101–120, 1931.

[8] Edwin P Hubble. *106. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae*. Harvard University Press, 2013.

[9] George Gamow. Expanding universe and the origin of elements. *Physical review*, 70(7-8): 572, 1946.

[10] Arno A Penzias and Robert Woodrow Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419–421, 1965.

[11] Nancy W Boggess, John C Mather, R Weiss, Charles L Bennett, ESe al Cheng, E Dwek, S Gulkis, Michael G Hauser, Michael A Janssen, T Kelsall, et al. The cobe mission-its design and performance two years after launch. *The Astrophysical Journal*, 397:420–429, 1992.

[12] Gary Hinshaw, D Larson, Eiichiro Komatsu, David N Spergel, CLaa Bennett, Joanna Dunkley, MR Nolta, M Halpern, RS Hill, N Odegard, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19, 2013.

[13] Yashar Akrami, M Ashdown, Jonathan Aumont, Carlo Baccigalupi, M Ballardini, Anthony J Banday, RB Barreiro, Nicola Bartolo, S Basak, K Benabed, et al. Planck 2018 results-iv. diffuse component separation. *Astronomy & Astrophysics*, 641:A4, 2020.

[14] DJ Fixsen, ES Cheng, JM Gales, John C Mather, RA Shafer, and EL Wright. The cosmic microwave background spectrum from the full cobe* firas data set. *The Astrophysical Journal*, 473(2):576, 1996.

[15] Alexei A Starobinsky. A new type of isotropic cosmological models without singularity. *Physics Letters B*, 91(1):99–102, 1980.

[16] Alan H Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347, 1981.

[17] Andrei D Linde. A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108 (6):389–393, 1982.

[18] Fritz Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica physica acta*, 6:110–127, 1933.

[19] Luca Amendola and Shinji Tsujikawa. *Dark Energy: Theory and Observations*. 2010.

[20] Adam G Riess, Alexei V Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M Garnavich, Ron L Gilliland, Craig J Hogan, Saurabh Jha, Robert P Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.

[21] Saul Perlmutter, Goldhaber Aldering, Gerson Goldhaber, RA Knop, Peter Nugent, Patricia G Castro, Susana Deustua, Sebastien Fabbro, Ariel Goobar, Donald E Groom, et al. Measurements of $\omega$ and $\lambda$ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999.

[22] Luca Amendola and Shinji Tsujikawa. *Dark energy: theory and observations*. Cambridge University Press, 2010.

[23] Volker Springel, Simon DM White, Adrian Jenkins, Carlos S Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature*, 435(7042):629–636, 2005.

[24] William H Press and Paul Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, 187:425–438, 1974.

[25] Jeremy Tinker, Andrey V Kravtsov, Anatoly Klypin, Kevork Abazajian, Michael Warren, Gustavo Yepes, Stefan Gottlöber, and Daniel E Holz. Toward a halo mass function for

precision cosmology: the limits of universality. *The Astrophysical Journal*, 688(2):709, 2008.

[26] Anatoly Klypin. Numerical simulations in cosmology. *Electron–Positron Physics at the Z*, page 420, 1996.

[27] Gustavo Yepes. The universe in a computer: The importance of numerical simulations in cosmology. In *Historical Development of Modern Cosmology*, volume 252, page 355, 2001.

[28] Sebastian Von Hoerner. Die numerische integration des n-körper-problemes für sternhaufen. i. *Zeitschrift fur Astrophysik*, 50, 1960.

[29] Sverre J Aarseth and F Hoyle. Dynamical evolution of clusters of galaxies, i. *Monthly Notices of the Royal Astronomical Society*, 126(3):223–255, 1963.

[30] PJE Peebles. Structure of the coma cluster of galaxies. *The Astronomical Journal*, 75:13, 1970.

[31] Simon DM White. The dynamics of rich clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, 177(3):717–733, 1976.

[32] Sverre J Aarseth, J Richard Gott III, and Edwin L Turner. N-body simulations of galaxy clustering. i-initial conditions and galaxy collapse times. *The Astrophysical Journal*, 228: 664–683, 1979.

[33] Ya B Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and astrophysics*, 5:84–89, 1970.

[34] AG Doroshkevich, EV Kotok, ID Novikov, AN Polyudov, SF Shandarin, and Yu S Sigov. Two-dimensional simulation of the gravitational system dynamics and formation of the large-scale structure of the universe. *Monthly Notices of the Royal Astronomical Society*, 192(2):321–337, 1980.

[35] AA Klypin and SF Shandarin. Three-dimensional numerical model of the formation of large-scale structure in the universe. *Monthly Notices of the Royal Astronomical Society*, 204(3):891–907, 1983.

[36] George Efstathiou and JW Eastwood. On the clustering of particles in an expanding universe. *Monthly Notices of the Royal Astronomical Society*, 194(3):503–525, 1981.

[37] George Efstathiou, Marc Davis, SDM White, and CS Frenk. Numerical techniques for large cosmological n-body simulations. *The Astrophysical Journal Supplement Series*, 57: 241–260, 1985.

[38] HMP Couchman. Mesh-refined p3m-a fast adaptive n-body algorithm. *The Astrophysical Journal*, 368:L23–L26, 1991.

[39] Josh Barnes and Piet Hut. A hierarchical o (n log n) force-calculation algorithm. *nature*, 324(6096):446–449, 1986.

[40] Wojciech H Zurek, Peter J Quinn, John K Salmon, and Michael S Warren. Large-scale structure after cobe: Peculiar velocities and correlations of cold dark matter halos. *The Astrophysical Journal*, 431:559–568, 1994.

[41] Rene Laureijs, J Amiaux, S Arduini, J-L Augueres, J Brinchmann, R Cole, M Cropper, C Dabin, L Duvet, A Ealet, et al. Euclid definition study report. *arXiv preprint arXiv:1110.3193*, 2011.

[42] Anatoly Klypin, Gustavo Yepes, Stefan Gottlöber, Francisco Prada, and Steffen Hess. Multidark simulations: the story of dark matter halo concentrations and density profiles. *Monthly Notices of the Royal Astronomical Society*, 457(4):4340–4359, 2016.

[43] Weiguang Cui, Alexander Knebe, Gustavo Yepes, Frazer Pearce, Chris Power, Romeel Dave, Alexander Arth, Stefano Borgani, Klaus Dolag, Pascal Elahi, et al. The three hundred project: a large catalogue of theoretically modelled galaxy clusters for cosmological and astrophysical applications. *Monthly Notices of the Royal Astronomical Society*, 480 (3):2898–2915, 2018.

[44] Volker Springel, Naoki Yoshida, and Simon DM White. Gadget: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6(2):79–117, 2001.

[45] G Kauffmann, B Guiderdoni, and SDM White. Faint galaxy counts in a hierarchical universe. *Monthly Notices of the Royal Astronomical Society*, 267(4):981–999, 1994.

[46] Shaun Cole, Alfonso Aragon-Salamanca, Carlos S Frenk, Julio F Navarro, and Stephen E Zepf. A recipe for galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 271(4):781–806, 1994.

[47] Rachel S Somerville and Joel R Primack. Semi-analytic modelling of galaxy formation: the local universe. *Monthly Notices of the Royal Astronomical Society*, 310(4):1087–1110, 1999.

[48] Guinevere Kauffmann, Adi Nusser, and Matthias Steinmetz. Galaxy formation and large-scale bias. *Monthly Notices of the Royal Astronomical Society*, 286(4):795–811, 1997.

[49] Julio F Navarro and Simon DM White. Simulations of dissipative galaxy formation in hierarchically clustering universes–i: Tests of the code. *Monthly Notices of the Royal Astronomical Society*, 265(2):271–300, 1993.

[50] G Yepes, R Kates, A Khokhlov, and A Klypin. Hydrodynamical simulations of galaxy formation: effects of supernova feedback. *Monthly Notices of the Royal Astronomical Society*, 284(1):235–256, 1997.

[51] Giovanni Carraro, Cesario Lia, and Cesare Chiosi. Galaxy formation and evolution-i. the padua tree-sph code (pd-sph). *Monthly Notices of the Royal Astronomical Society*, 297(4): 1021–1040, 1998.

[52] Begoña Ascaso, Simona Mei, and Narciso Benítez. Apples to apples a2–i. realistic galaxy simulated catalogues and photometric redshift predictions for next-generation surveys. *Monthly Notices of the Royal Astronomical Society*, 453(3):2515–2532, 2015.

[53] Paul A Abell, Julius Allison, Scott F Anderson, John R Andrew, J Roger P Angel, Lee Armus, David Arnett, SJ Asztalos, Tim S Axelrod, Stephen Bailey, et al. Lsst science book, version 2.0. *arXiv preprint arXiv:0912.0201*, 2009.

[54] Leon B Lucy. A numerical approach to the testing of the fission hypothesis. *The astronomical journal*, 82:1013–1024, 1977.

[55] Robert A Gingold and Joseph J Monaghan. Smoothed particle hydrodynamics: theory and application to non-spherical stars. *Monthly notices of the royal astronomical society*, 181(3):375–389, 1977.

[56] August E Evrard. Beyond n-body-3d cosmological gas dynamics. *Monthly Notices of the Royal Astronomical Society*, 235:911–934, 1988.

[57] Lars Hernquist and Neal Katz. TREESPH: A Unification of SPH with the Hierarchical Tree Method. , 70:419, June 1989. doi: 10.1086/191344.

[58] Lars Hernquist and Neal Katz. Treesph-a unification of sph with the hierarchical tree method. *The Astrophysical Journal Supplement Series*, 70:419–446, 1989.

[59] Volker Springel. The cosmological simulation code gadget-2. *Monthly notices of the royal astronomical society*, 364(4):1105–1134, 2005.

[60] RY Cen, Antony Jameson, Feng Liu, and Jeremiah P Ostriker. The universe in a box-thermal effects in the standard cold dark matter scenario. *The Astrophysical Journal*, 362: L41–L45, 1990.

[61] Sergei Konstantinovich Godunov and Viktor Solomonovich Ryabenkii. *Difference schemes: an introduction to the underlying theory*. Elsevier, 1987.

[62] Phillip Colella and Paul R Woodward. The piecewise parabolic method (ppm) for gas-dynamical simulations. *Journal of computational physics*, 54(1):174–201, 1984.

[63] Greg L Bryan, Renyue Cen, Michael L Norman, Jeremiah P Ostriker, and James M Stone. X-ray clusters from a high-resolution hydrodynamic ppm simulation of the cold dark matter universe. *The Astrophysical Journal*, 428:405–418, 1994.

[64] Vicent Quilis, Jose M. A. Ibanez, and Diego Saez. A Multidimensional Hydrodynamic Code for Structure Evolution in Cosmology. , 469:11, September 1996. doi: 10.1086/177753.

[65] Andrew Sornborger, Robert Brandenberger, Bruce Fryxell, and Kevin Olson. The structure of cosmic string wakes. *The Astrophysical Journal*, 482(1):22, 1997.

[66] Dongsu Ryu, Jeremiah P. Ostriker, Hyesung Kang, and Renyue Cen. A Cosmological Hydrodynamic Code Based on the Total Variation Diminishing Scheme. , 414:1, September 1993. doi: 10.1086/173051.

[67] A. Klypin, R. Kates, and A. Khokhlov. Galaxy formation with gravitation, hydrodynamics and active star formation. In Vincent J. Martínez, Miquel Portilla, and Diego Sáez, editors, *New Insights into the Universe*, pages 157–184, Berlin, Heidelberg, 1992. Springer Berlin Heidelberg. ISBN 978-3-540-47296-4.

[68] S. Maurogordato, C. Balkowski, C. Tao, and J. Tran Thanh Van, editors. *Clustering in the universe. Proceedings, 30th Rencontres de Moriond, 15th Moriond Astrophysics Meetings, Les Arcs, France, March 11-18, 1995*, Gif-Sur-Yvette, 1995. Ed. Frontieres.

[69] Kurt Roettiger, Jack Burns, and Chris Loken. When clusters collide-a numerical hydro/n-body simulation of merging galaxy clusters. *The Astrophysical Journal*, 407:L53–L56, 1993.

[70] Peter Anninos, Michael L Norman, and David A Clarke. Hierarchical numerical cosmology with hydrodynamics: Methods and code tests. *The Astrophysical Journal*, 436:11–22, 1994.

[71] Peter Anninos, Michael L. Norman, and David A. Clarke. Hierarchical Numerical Cosmology with Hydrodynamics: Methods and Code Tests. , 436:11, November 1994. doi: 10.1086/174876.

[72] L Perotto, N Ponthieu, JF Macías-Pérez, R Adam, P Ade, P André, A Andrianasolo, H Aussel, A Beelen, A Benoît, et al. Calibration and performance of the nika2 camera at the iram 30-m telescope. *Astronomy & Astrophysics*, 637:A71, 2020.

[73] Peter AR Ade, N Aghanim, F Argüeso, C Armitage-Caplan, M Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, Carlo Baccigalupi, AJ Banday, et al. Planck 2013 results. xxviii. the planck catalogue of compact sources. *Astronomy & Astrophysics*, 571:A28, 2014.

[74] M. Hilton, C. Sifón, S. Naess, M. Madhavacheril, M. Oguri, E. Rozo, E. Rykoff, T. M. C. Abbott, S. Adhikari, M. Aguena, S. Aiola, S. Allam, S. Amodeo, A. Amon, J. Annis, B. Ansarinejad, C. Aros-Bunster, J. E. Austermann, S. Avila, D. Bacon, N. Battaglia, J. A. Beall, D. T. Becker, G. M. Bernstein, E. Bertin, T. Bhandarkar, S. Bhargava, J. R. Bond, D. Brooks, D. L. Burke, E. Calabrese, M. Carrasco Kind, J. Carretero, S. K. Choi, A. Choi, C. Conselice, L. N. da Costa, M. Costanzi, D. Crichton, K. T. Crowley, R. Dünner, E. V. Denison, M. J. Devlin, S. R. Dicker, H. T. Diehl, J. P. Dietrich, P. Doel, S. M. Duff, A. J. Duivenvoorden, J. Dunkley, S. Everett, S. Ferraro, I. Ferrero, A. Ferté, B. Flaugher, J. Frieman, P. A. Gallardo, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, P. Giles, J. E. Golec, M. B. Gralla, S. Grandis, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. Han, W. G. Hartley, M. Hasselfield, J. C. Hill, G. C. Hilton, A. D. Hincks, S. R. Hinton, S. P. P. Ho, K. Honscheid, B. Hoyle, J. Hubmayr, K. M. Huffenberger, J. P. Hughes, A. T. Jaelani, B. Jain, D. J. James, T. Jeltema, S. Kent, K. Knowles, B. J. Koopman, K. Kuehn, O. Lahav, M. Lima, Y. T. Lin, M. Lokken, S. I. Loubser, N. MacCrann, M. A. G. Maia, T. A. Marriage, J. Martin, J. McMahon, P. Melchior, F. Menanteau, R. Miquel, H. Miyatake, K. Moodley, R. Morgan, T. Mroczkowski, F. Nati, L. B. Newburgh, M. D. Niemack, A. J. Nishizawa, R. L. C. Ogando, J. Orlowski-Scherer, L. A. Page, A. Palmese, B. Partridge, F. Paz-Chinchón, P. Phakathi, A. A. Plazas, N. C. Robertson, A. K. Romer, A. Carnero Rosell, M. Salatino, E. Sanchez, E. Schaan, A. Schillaci, N. Sehgal, S. Serrano, T. Shin, S. M. Simon, M. Smith, M. Soares-Santos, D. N. Spergel, S. T. Staggs, E. R. Storer, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, H. Trac, J. N. Ullom, L. R. Vale, J. Van Lanen, E. M. Vavagiakis, J. De Vicente, R. D. Wilkinson, E. J. Wollack, Z. Xu, and Y. Zhang. The Atacama Cosmology Telescope: A Catalog of >4000 Sunyaev–Zel'dovich Galaxy Clusters. , 253(1):3, March 2021. doi: 10.3847/1538-4365/abd023.

[75] L. E. Bleem, B. Stalder, T. de Haan, K. A. Aird, S. W. Allen, D. E. Applegate, M. L. N. Ashby, M. Bautz, M. Bayliss, B. A. Benson, S. Bocquet, M. Brodwin, J. E. Carlstrom, C. L. Chang, I. Chiu, H. M. Cho, A. Clocchiatti, T. M. Crawford, A. T. Crites, S. Desai, J. P. Dietrich, M. A. Dobbs, R. J. Foley, W. R. Forman, E. M. George, M. D. Gladders, A. H. Gonzalez, N. W. Halverson, C. Hennig, H. Hoekstra, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, C. Jones, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, J. Liu, M. Lueker, D. Luong-Van, A. Mantz, D. P. Marrone, M. McDonald, J. J. McMahon, S. S. Meyer, L. Mocanu, J. J. Mohr, S. S. Murray, S. Padin, C. Pryke, C. L. Reichardt, A. Rest, J. Ruel, J. E. Ruhl, B. R. Saliwanchik, A. Saro, J. T. Sayre, K. K. Schaffer, T. Schrabback, E. Shirokoff, J. Song, H. G. Spieler, S. A. Stanford, Z. Staniszewski, A. A. Stark, K. T. Story, C. W. Stubbs, K. Vanderlinde, J. D. Vieira, A. Vikhlinin, R. Williamson, O. Zahn, and A. Zenteno. Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree

SPT-SZ Survey. , 216(2):27, February 2015. doi: 10.1088/0067-0049/216/2/27.

[76] RA Sunyaev and Ya B Zeldovich. The observations of relic radiation as a test of the nature of x-ray radiation from the clusters of galaxies. *Comments on Astrophysics and Space Physics*, 4:173, 1972.

[77] RA Sunyaev and Ya B Zel'Dovich. Microwave background radiation as a probe of the contemporary structure and history of the universe. *Annual review of astronomy and astrophysics*, 18(1):537–560, 1980.

[78] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, B. Comis, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, M. Giard, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, F. K. Hansen, D. Hanson, D. L. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, R. Leonardi, J. Lesgourgues, F. Levrier, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, M. Maris, P. G. Martin, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, R. Rebolo, M. Reinecke, M. Re-

mazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, M. Roman, C. Rosset, M. Rossetti, G. Roudier, J. A. Rubiño-Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, J. Weller, S. D. M. White, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts. , 594:A24, September 2016. doi: 10.1051/0004-6361/201525833.

[79] S. Bocquet, J. P. Dietrich, T. Schrabback, L. E. Bleem, M. Klein, S. W. Allen, D. E. Applegate, M. L. N. Ashby, M. Bautz, M. Bayliss, B. A. Benson, M. Brodwin, E. Bulbul, R. E. A. Canning, R. Capasso, J. E. Carlstrom, C. L. Chang, I. Chiu, H. M. Cho, A. Clocchiatti, T. M. Crawford, A. T. Crites, T. de Haan, S. Desai, M. A. Dobbs, R. J. Foley, W. R. Forman, G. P. Garmire, E. M. George, M. D. Gladders, A. H. Gonzalez, S. Grandis, N. Gupta, N. W. Halverson, J. Hlavacek-Larrondo, H. Hoekstra, G. P. Holder, W. L. Holzapfel, Z. Hou, J. D. Hrubes, N. Huang, C. Jones, G. Khullar, L. Knox, R. Kraft, A. T. Lee, A. von der Linden, D. Luong-Van, A. Mantz, D. P. Marrone, M. McDonald, J. J. McMahon, S. S. Meyer, L. M. Mocanu, J. J. Mohr, R. G. Morris, S. Padin, S. Patil, C. Pryke, D. Rapetti, C. L. Reichardt, A. Rest, J. E. Ruhl, B. R. Saliwanchik, A. Saro, J. T. Sayre, K. K. Schaffer, E. Shirokoff, B. Stalder, S. A. Stanford, Z. Staniszewski, A. A. Stark, K. T. Story, V. Strazzullo, C. W. Stubbs, K. Vanderlinde, J. D. Vieira, A. Vikhlinin, R. Williamson, and A. Zenteno. Cluster Cosmology Constraints from the 2500 deg$^2$ SPT-SZ Survey: Inclusion of Weak Gravitational Lensing Data from Magellan and the Hubble Space Telescope. , 878(1):55, June 2019. doi: 10.3847/1538-4357/ab1f10.

[80] Matthew Hasselfield, Matt Hilton, Tobias A. Marriage, Graeme E. Addison, L. Felipe Barrientos, Nicholas Battaglia, Elia S. Battistelli, J. Richard Bond, Devin Crichton, Sudeep Das, Mark J. Devlin, Simon R. Dicker, Joanna Dunkley, Rolando Dünner, Joseph W. Fowler, Megan B. Gralla, Amir Hajian, Mark Halpern, Adam D. Hincks, Renée Hlozek, John P. Hughes, Leopoldo Infante, Kent D. Irwin, Arthur Kosowsky, Danica Marsden, Felipe Menanteau, Kavilan Moodley, Michael D. Niemack, Michael R. Nolta, Lyman A. Page, Bruce Partridge, Erik D. Reese, Benjamin L. Schmitt, Neelima Sehgal, Blake D. Sherwin, Jon Sievers, Cristóbal Sifón, David N. Spergel, Suzanne T. Staggs, Daniel S. Swetz, Eric R. Switzer, Robert Thornton, Hy Trac, and Edward J. Wollack. The Atacama Cosmology Telescope: Sunyaev-Zel'dovich selected galaxy clusters at 148 GHz from three seasons of data. , 2013(7):008, July 2013. doi: 10.1088/1475-7516/2013/07/008.

[81] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara,

J. Annis, J. Asorey, S. Avila, D. Bacon, E. Balbinot, M. Banerji, N. Banik, W. Barkhouse, M. Baumer, E. Baxter, K. Bechtol, M. R. Becker, A. Benoit-Lévy, B. A. Benson, G. M. Bernstein, E. Bertin, J. Blazek, S. L. Bridle, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, M. T. Busha, A. Campos, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, N. Chen, M. Childress, A. Choi, C. Conselice, R. Crittenden, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, R. Das, T. M. Davis, C. Davis, J. De Vicente, D. L. DePoy, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, A. Drlica-Wagner, T. F. Eifler, A. E. Elliott, F. Elsner, J. Elvin-Poole, J. Estrada, A. E. Evrard, Y. Fang, E. Fernandez, A. Ferté, D. A. Finley, B. Flaugher, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Garcia-Fernandez, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, M. S. S. Gill, K. Glazebrook, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. Hamilton, W. G. Hartley, S. R. Hinton, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, T. Kacprzak, S. Kent, A. G. Kim, A. King, D. Kirk, N. Kokron, A. Kovacs, E. Krause, C. Krawiec, A. Kremin, K. Kuehn, S. Kuhlmann, N. Kuropatkin, F. Lacasa, O. Lahav, T. S. Li, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, M. Makler, M. Manera, M. March, J. L. Marshall, P. Martini, R. G. McMahon, P. Melchior, F. Menanteau, R. Miquel, V. Miranda, D. Mudd, J. Muir, A. Möller, E. Neilsen, R. C. Nichol, B. Nord, P. Nugent, R. L. C. Ogando, A. Palmese, J. Peacock, H. V. Peiris, J. Peoples, W. J. Percival, D. Petravick, A. A. Plazas, A. Porredon, J. Prat, A. Pujol, M. M. Rau, A. Refregier, P. M. Ricker, N. Roe, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. Rozo, E. S. Rykoff, M. Sako, A. I. Salvador, S. Samuroff, C. Sánchez, E. Sanchez, B. Santiago, V. Scarpine, R. Schindler, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, R. C. Smith, M. Smith, J. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, D. L. Tucker, B. E. Tucker, S. A. Uddin, T. N. Varga, P. Vielzeuf, V. Vikram, A. K. Vivas, A. R. Walker, M. Wang, R. H. Wechsler, J. Weller, W. Wester, R. C. Wolf, B. Yanny, F. Yuan, A. Zenteno, B. Zhang, Y. Zhang, J. Zuntz, and Dark Energy Survey Collaboration. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. , 98(4):043526, August 2018. doi: 10.1103/PhysRevD.98.043526.

[82] F Jansen, D Lumb, B Altieri, J Clavel, M Ehle, C Erd, C Gabriel, M Guainazzi, P Gondoin, R Much, et al. Xmm-newton observatory-i. the spacecraft and operations. *Astronomy & Astrophysics*, 365(1):L1–L6, 2001.

[83] Annalisa Pillepich, Thomas H. Reiprich, Cristiano Porciani, Katharina Borm, and Andrea Merloni. Forecasts on dark energy from the X-ray cluster survey with eROSITA:

constraints from counts and clustering. , 481(1):613–626, November 2018. doi: 10.1093/mnras/sty2240.

[84] M. C. Weisskopf, T. L. Aldcroft, M. Bautz, R. A. Cameron, D. Dewey, J. J. Drake, C. E. Grant, H. L. Marshall, and S. S. Murray. An Overview of the Performance of the Chandra X-ray Observatory. *Experimental Astronomy*, 16(1):1–68, August 2003. doi: 10.1023/B:EXPA.0000038953.49421.54.

[85] D Fabjan, Stefano Borgani, E Rasia, A Bonafede, K Dolag, G Murante, and Luca Tornatore. X-ray mass proxies from hydrodynamic simulations of galaxy clusters–i. *Monthly Notices of the Royal Astronomical Society*, 416(2):801–816, 2011.

[86] Andrea Biviano. Galaxy systems in the optical and infrared. *arXiv preprint arXiv:0811.3535*, 2008.

[87] G. F. Lesci, F. Marulli, L. Moscardini, M. Sereno, A. Veropalumbo, M. Maturi, C. Giocoli, M. Radovich, F. Bellagamba, M. Roncarelli, S. Bardelli, S. Contarini, G. Covone, L. Ingoglia, L. Nanni, and E. Puddu. AMICO galaxy clusters in KiDS-DR3: cosmological constraints from counts and stacked weak-lensing. *arXiv e-prints*, art. arXiv:2012.12273, December 2020.

[88] M. Costanzi, E. Rozo, M. Simet, Y. Zhang, A. E. Evrard, A. Mantz, E. S. Rykoff, T. Jeltema, D. Gruen, S. Allen, T. McClintock, A. K. Romer, A. von der Linden, A. Farahi, J. DeRose, T. N. Varga, J. Weller, P. Giles, D. L. Hollowood, S. Bhargava, A. Bermeo-Hernandez, X. Chen, T. M. C. Abbott, F. B. Abdalla, S. Avila, K. Bechtol, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, M. Crocce, C. E. Cunha, L. N. da Costa, C. Davis, J. De Vicente, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, J. Estrada, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, K. Honscheid, B. Hoyle, D. J. James, E. Krause, K. Kuehn, N. Kuropatkin, M. Lima, H. Lin, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, F. Menanteau, C. J. Miller, R. Miquel, J. J. Mohr, R. L. C. Ogando, A. A. Plazas, A. Roodman, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, M. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, and R. H. Wechsler. Methods for cluster cosmology and application to the SDSS in preparation for DES Year 1 release. , 488(4):4779–4800, October 2019. doi: 10.1093/mnras/stz1949.

[89] S. Andreon and J. Bergé. Richness-mass relation self-calibration for galaxy clusters. , 547: A117, November 2012. doi: 10.1051/0004-6361/201220115.

[90] J Binney and S Tremaine. Galaxy dynamics (princeton, nj, 1987.

[91] RG Carlberg, HKCr Yee, and E Ellingson. The average mass and light profiles of galaxy clusters. *The Astrophysical Journal*, 478(2):462, 1997.

[92] S. Andreon and M. A. Hurn. The scaling relation between richness and mass of galaxy clusters: a Bayesian approach. , 404(4):1922–1937, June 2010. doi: 10.1111/j.1365-2966.2010.16406.x.

[93] Paul Schechter. An analytic expression for the luminosity function for galaxies. *The Astrophysical Journal*, 203:297–306, 1976.

[94] John Beverley Oke. Absolute spectral energy distributions for white dwarfs. *The Astrophysical Journal Supplement Series*, 27:21, 1974.

[95] Bianca Maria Poggianti. K and evolutionary corrections from uv to ir. *Astronomy and Astrophysics Supplement Series*, 122(3):399–407, 1997.

[96] David W Hogg, Ivan K Baldry, Michael R Blanton, and Daniel J Eisenstein. The k correction. *arXiv preprint astro-ph/0210394*, 2002.

[97] SP Driver, S Phillipps, Jonathan Ivor Davies, I Morgan, and Michael John Disney. Dwarf galaxies at: photometry of the cluster abell 963. *Monthly Notices of the Royal Astronomical Society*, 268(2):393–404, 1994.

[98] Andrea V Macciò, Aaron A Dutton, and Frank C Van Den Bosch. Concentration, spin and shape of dark matter haloes as a function of the cosmological model: Wmap 1, wmap 3 and wmap 5 results. *Monthly Notices of the Royal Astronomical Society*, 391(4):1940–1954, 2008.

[99] Liang Gao, Gabriella De Lucia, Simon DM White, and Adrian Jenkins. Galaxies and subhaloes in $\lambda$cdm galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 352(2):L1–L5, 2004.

[100] Darren Reed, Fabio Governato, Licia Verde, Jeffrey Gardner, Thomas Quinn, Joachim Stadel, David Merritt, and George Lake. Evolution of the density profiles of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 357(1):82–96, 2005.

[101] Julio F Navarro. The structure of cold dark matter halos. In *Symposium-international astronomical union*, volume 171, pages 255–258. Cambridge University Press, 1996.

[102] Aaron D Ludlow, Julio F Navarro, Raúl E Angulo, Michael Boylan-Kolchin, Volker Springel, Carlos Frenk, and Simon DM White. The mass–concentration–redshift relation of cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 441 (1):378–388, 2014.

[103] Liang Gao, Julio F Navarro, Shaun Cole, Carlos S Frenk, Simon DM White, Volker Springel, Adrian Jenkins, and Angelo F Neto. The redshift dependence of the structure of massive $\lambda$ cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 387(2):536–544, 2008.

[104] Jaan Einasto. On the construction of a composite model for the galaxy and on the determination of the system of galactic parameters. *Trudy Astrofizicheskogo Instituta Alma-Ata*, 5:87–100, 1965.

[105] Julio F Navarro, Eric Hayashi, Chris Power, AR Jenkins, Carlos S Frenk, Simon DM White, Volker Springel, Joachim Stadel, and Thomas R Quinn. The inner structure of $\lambda$cdm haloes–iii. universality and asymptotic slopes. *Monthly Notices of the Royal Astronomical Society*, 349(3):1039–1051, 2004.

[106] Julio F Navarro, Aaron Ludlow, Volker Springel, Jie Wang, Mark Vogelsberger, Simon DM White, Adrian Jenkins, Carlos S Frenk, and Amina Helmi. The diversity and similarity of simulated cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 402(1):21–34, 2010.

[107] WJG De Blok, Fabian Walter, Elias Brinks, C Trachternach, SH Oh, and RC Kennicutt Jr. High-resolution rotation curves and galaxy mass models from things. *The Astronomical Journal*, 136(6):2648, 2008.

[108] Andreas Burkert. The structure of dark matter halos in dwarf galaxies. *The Astrophysical Journal Letters*, 447(1):L25, 1995.

[109] Yinon Arieli and Yoel Rephaeli. Dark matter profiles in clusters of galaxies: a phenomenological approach. *New Astronomy*, 8(6):517–528, 2003.

[110] Tereasa G Brainerd and Candace Oaxaca Wright. Constraining galaxy halo shapes with weak lensing. In *A New Era in Cosmology*, volume 283, page 177, 2002.

[111] Fabrice Brimioulle, S Seitz, M Lerchster, R Bender, and J Snigula. Dark matter halo properties from galaxy–galaxy lensing. *Monthly Notices of the Royal Astronomical Society*, 432(2):1046–1102, 2013.

[112] Andrea Biviano. From messier to abell: 200 years of science with galaxy clusters. *arXiv preprint astro-ph/0010409*, 2000.

[113] George O Abell. The distribution of rich clusters of galaxies. *The Astrophysical Journal Supplement Series*, 3:211, 1958.

[114] Marc Postman, Lori M Lubin, James E Gunn, JB Oke, John G Hoessel, Donald P Schneider, and Jennifer A Christensen. The palomar distant cluster survey: I. the cluster catalog. *arXiv preprint astro-ph/9511011*, 1995.

[115] Neta A Bahcall, Timothy A McKay, James Annis, Rita SJ Kim, Feng Dong, Sarah Hansen, Tomo Goto, James E Gunn, Chris Miller, Robert C Nichol, et al. A merged catalog of clusters of galaxies from early sloan digital sky survey data. *The Astrophysical Journal Supplement Series*, 148(2):243, 2003.

[116] LF Olsen, M Scodeggio, L Da Costa, C Benoist, E Bertin, E Deul, T Erben, MD Guarnieri, R Hook, M Nonino, et al. Eso imaging survey ii. searching for distant clusters of galaxies. *arXiv preprint astro-ph/9803338*, 1998.

[117] CS Kochanek, Martin White, J Huchra, L Macri, TH Jarrett, SE Schneider, and J Mader. Clusters of galaxies in the local universe. *The Astrophysical Journal*, 585(1):161, 2003.

[118] Jeffrey A Willick, Keith L Thompson, Benjamin F Mathiesen, Saul Perlmutter, Robert A Knop, and Gary J Hill. The stanford cluster search: Scope, method, and preliminary results. *Publications of the Astronomical Society of the Pacific*, 113(784):658, 2001.

[119] JP Huchra and MJs Geller. Groups of galaxies. i-nearby groups. *The Astrophysical Journal*, 257:423–437, 1982.

[120] MJ Geller and JP Huchra. Groups of galaxies. iii-the cfa survey. *The Astrophysical Journal Supplement Series*, 52:61–87, 1983.

[121] Vincent R Eke, Carlton M Baugh, Shaun Cole, Carlos S Frenk, Peder Norberg, John A Peacock, Ivan K Baldry, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, et al. Galaxy groups in the 2dfgrs: the group-finding algorithm and the 2pigg catalogue. *Monthly Notices of the Royal Astronomical Society*, 348(3):866–878, 2004.

[122] Andreas A Berlind, Joshua Frieman, David H Weinberg, Michael R Blanton, Michael S Warren, Kevork Abazajian, Ryan Scranton, David W Hogg, Roman Scoccimarro, Neta A Bahcall, et al. Percolation galaxy groups and clusters in the sdss redshift survey: identification, catalogs, and the multiplicity function. *The Astrophysical Journal Supplement Series*, 167(1):1, 2006.

[123] Manuel E Merchán and Ariel Zandivarez. Galaxy groups in the third data release of the sloan digital sky survey. *The Astrophysical Journal*, 630(2):759, 2005.

[124] Massimo Ramella, Armando Pisani, and Margaret J Geller. Groups of galaxies in the northern cfa redshift survey. *The Astronomical Journal*, 113:483, 1997.

[125] Aidan C Crook, John P Huchra, Nathalie Martimbeau, Karen L Masters, Tom Jarrett, and Lucas M Macri. Groups of galaxies in the two micron all sky redshift survey. *The Astrophysical Journal*, 655(2):790, 2007.

[126] Michael D Gladders and HKC Yee. A new method for galaxy cluster detection. i. the algorithm. *The Astronomical Journal*, 120(4):2148, 2000.

[127] Peter RM Eisenhardt, Mark Brodwin, Anthony H Gonzalez, S Adam Stanford, Daniel Stern, Pauline Barmby, Michael JI Brown, Kyle Dawson, Arjun Dey, Mamoru Doi, et al. Clusters of galaxies in the first half of the universe from the irac shallow survey. *The Astrophysical Journal*, 684(2):905, 2008.

[128] Gillian Wilson, Adam Muzzin, Mark Lacy, Howard Yee, Jason Surace, Carol Lonsdale, Henk Hoekstra, Subhabrata Majumdar, David Gilbank, and Mike Gladders. Clusters of galaxies at 1< z< 2: the spitzer adaptation of the red-sequence cluster survey. *arXiv preprint astro-ph/0604289*, 2006.

[129] Michael D Gladders and Howard KC Yee. The red-sequence cluster survey. i. the survey and cluster catalogs for patches rcs 0926+ 37 and rcs 1327+ 29. *The Astrophysical Journal Supplement Series*, 157(1):1, 2005.

[130] Michael D Gladders, HKC Yee, Subhabrata Majumdar, L Felipe Barrientos, Henk Hoekstra, Patrick B Hall, and Leopoldo Infante. Cosmological constraints from the red-sequence cluster survey. *The Astrophysical Journal*, 655(1):128, 2007.

[131] Benjamin P Koester, Timothy A McKay, James Annis, Risa H Wechsler, August E Evrard, Eduardo Rozo, Lindsey Bleem, Erin S Sheldon, and David Johnston. Maxbcg: A red-sequence galaxy cluster finder. *The Astrophysical Journal*, 660(1):221, 2007.

[132] BP Koester, Timothy A McKay, James Annis, Risa H Wechsler, A Evrard, L Bleem, M Becker, D Johnston, E Sheldon, R Nichol, et al. A maxbcg catalog of 13,823 galaxy clusters from the sloan digital sky survey. *The Astrophysical Journal*, 660(1):239, 2007.

[133] Fabio Bellagamba, Mauro Roncarelli, Matteo Maturi, and Lauro Moscardini. Amico: optimized detection of galaxy clusters in photometric surveys. *Monthly Notices of the Royal Astronomical Society*, 473(4):5221–5236, 2018.

[134] R Adam, M Vannier, S Maurogordato, ANDREA Biviano, C Adami, B Ascaso, F Bellagamba, C Benoist, Alberto Cappi, A Díaz-Sánchez, et al. Euclid preparation-iii. galaxy

cluster detection in the wide photometric survey, performance and algorithm selection. *Astronomy & Astrophysics*, 627:A23, 2019.

[135] Matteo Maturi, Fabio Bellagamba, Mario Radovich, Mauro Roncarelli, Mauro Sereno, Lauro Moscardini, Sandro Bardelli, and Emanuella Puddu. AMICO galaxy clusters in KiDS-DR3: sample properties and selection function. , 485(1):498–512, May 2019. doi: 10.1093/mnras/stz294.

[136] Stefano Borgani and Luigi Guzzo. X-ray clusters of galaxies as tracers of structure in the universe. *Nature*, 409(6816):39–45, 2001.

[137] B. Sartoris, A. Biviano, C. Fedeli, J. G. Bartlett, S. Borgani, M. Costanzi, C. Giocoli, L. Moscardini, J. Weller, B. Ascaso, S. Bardelli, S. Maurogordato, and P. T. P. Viana. Next generation cosmology: constraints from the Euclid galaxy cluster survey. , 459(2): 1764–1780, June 2016. doi: 10.1093/mnras/stw630.

[138] E. S. Rykoff, E. Rozo, M. T. Busha, C. E. Cunha, A. Finoguenov, A. Evrard, J. Hao, B. P. Koester, A. Leauthaud, B. Nord, M. Pierre, R. Reddick, T. Sadibekova, E. S. Sheldon, and R. H. Wechsler. redMaPPer. I. Algorithm and SDSS DR8 Catalog. , 785(2):104, April 2014. doi: 10.1088/0004-637X/785/2/104.

[139] Steven W Allen, August E Evrard, and Adam B Mantz. Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 49:409–470, 2011.

[140] Boris Bolliet, Barbara Comis, Eiichiro Komatsu, and Juan Francisco Macías-Pérez. Dark energy constraints from the thermal Sunyaev-Zeldovich power spectrum. , 477(4):4957–4967, July 2018. doi: 10.1093/mnras/sty823.

[141] Stefano Ettori, Annamaria Donnarumma, Etienne Pointecouteau, Thomas H Reiprich, Stefania Giodini, Lorenzo Lovisari, and Robert W Schmidt. Mass profiles of galaxy clusters from x-ray analysis. *Space Science Reviews*, 177(1-4):119–154, 2013.

[142] Keiichi Umetsu. Cluster–galaxy weak lensing. *The Astronomy and Astrophysics Review*, 28(1):1–106, 2020.

[143] Eduardo Rozo, Eli S Rykoff, Benjamin P Koester, Timothy McKay, Jiangang Hao, August Evrard, Risa H Wechsler, Sarah Hansen, Erin Sheldon, David Johnston, et al. Improvement of the richness estimates of maxbcg clusters. *The Astrophysical Journal*, 703(1):601, 2009.

[144] SW Allen, DA Rapetti, RW Schmidt, H Ebeling, RG Morris, and AC Fabian. Improved constraints on dark energy from chandra x-ray observations of the largest relaxed galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 383(3):879–896, 2008.

[145] Stefano Ettori, A Morandi, P Tozzi, I Balestra, Stefano Borgani, Piero Rosati, L Lovisari, and F Terenziani. The cluster gas mass fraction as a cosmological probe: a revised study. *Astronomy & Astrophysics*, 501(1):61–73, 2009.

[146] Juan Estrada, Emiliano Sefusatti, and Joshua A Frieman. The correlation function of optically selected galaxy clusters in the sloan digital sky survey. *The Astrophysical Journal*, 692(1):265, 2009.

[147] Peter Schuecker, Hans Böhringer, Chris A Collins, and Luigi Guzzo. The reflex galaxy cluster survey-vii. and from cluster abundance and large-scale clustering. *Astronomy & Astrophysics*, 398(3):867–877, 2003.

[148] Annika HG Peter, Christopher E Moody, and Marc Kamionkowski. Dark-matter decays and self-gravitating halos. *Physical Review D*, 81(10):103501, 2010.

[149] L. Balkenhol, D. Dutcher, P. A. R. Ade, Z. Ahmed, E. Anderes, A. J. Anderson, M. Archipley, J. S. Avva, K. Aylor, P. S. Barry, R. Basu Thakur, K. Benabed, A. N. Bender, B. A. Benson, F. Bianchini, L. E. Bleem, F. R. Bouchet, L. Bryant, K. Byrum, J. E. Carlstrom, F. W. Carter, T. W. Cecil, C. L. Chang, P. Chaubal, G. Chen, H. M. Cho, T. L. Chou, J. F. Cliche, T. M. Crawford, A. Cukierman, C. Daley, T. de Haan, E. V. Denison, K. Dibert, J. Ding, M. A. Dobbs, W. Everett, C. Feng, K. R. Ferguson, A. Foster, J. Fu, S. Galli, A. E. Gambrel, R. W. Gardner, N. Goeckner-Wald, R. Gualtieri, S. Guns, N. Gupta, R. Guyser, N. W. Halverson, A. H. Harke-Hosemann, N. L. Harrington, J. W. Henning, G. C. Hilton, E. Hivon, G. P. Holder, W. L. Holzapfel, J. C. Hood, D. Howe, N. Huang, K. D. Irwin, O. B. Jeong, M. Jonas, A. Jones, T. S. Khaire, L. Knox, A. M. Kofman, M. Korman, D. L. Kubik, S. Kuhlmann, C. L. Kuo, A. T. Lee, E. M. Leitch, A. E. Lowitz, C. Lu, S. S. Meyer, D. Michalik, M. Millea, J. Montgomery, A. Nadolski, T. Natoli, H. Nguyen, G. I. Noble, V. Novosad, Y. Omori, S. Padin, Z. Pan, P. Paschos, J. Pearson, C. M. Posada, K. Prabhu, W. Quan, A. Rahlin, C. L. Reichardt, D. Riebel, B. Riedel, M. Rouble, J. E. Ruhl, J. T. Sayre, E. Schiappucci, E. Shirokoff, G. Smecher, J. A. Sobrin, A. A. Stark, J. Stephen, K. T. Story, A. Suzuki, K. L. Thompson, B. Thorne, C. Tucker, C. Umilta, L. R. Vale, K. Vanderlinde, J. D. Vieira, G. Wang, N. Whitehorn, W. L. K. Wu, V. Yefremenko, K. W. Yoon, and M. R. Young. Constraints on ΛCDM Extensions from the SPT-3G 2018 $EE$ and $TE$ Power Spectra. *arXiv e-prints*, art. arXiv:2103.13618, March 2021.

[150] Simone Aiola, Erminia Calabrese, Loïc Maurin, Sigurd Naess, Benjamin L. Schmitt, Maximilian H. Abitbol, Graeme E. Addison, Peter A. R. Ade, David Alonso, Mandana Amiri, Stefania Amodeo, Elio Angile, Jason E. Austermann, Taylor Baildon, Nick Battaglia, James A. Beall, Rachel Bean, Daniel T. Becker, J. Richard Bond, Sarah Marie Bruno, Victoria Calafut, Luis E. Campusano, Felipe Carrero, Grace E. Chesmore, Hsiao-mei Cho,

Steve K. Choi, Susan E. Clark, Nicholas F. Cothard, Devin Crichton, Kevin T. Crowley, Omar Darwish, Rahul Datta, Edward V. Denison, Mark J. Devlin, Cody J. Duell, Shannon M. Duff, Adriaan J. Duivenvoorden, Jo Dunkley, Rolando Dünner, Thomas Essinger-Hileman, Max Fankhanel, Simone Ferraro, Anna E. Fox, Brittany Fuzia, Patricio A. Gallardo, Vera Gluscevic, Joseph E. Golec, Emily Grace, Megan Gralla, Yilun Guan, Kirsten Hall, Mark Halpern, Dongwon Han, Peter Hargrave, Matthew Hasselfield, Jakob M. Helton, Shawn Henderson, Brandon Hensley, J. Colin Hill, Gene C. Hilton, Matt Hilton, Adam D. Hincks, Renée Hložek, Shuay-Pwu Patty Ho, Johannes Hubmayr, Kevin M. Huffenberger, John P. Hughes, Leopoldo Infante, Kent Irwin, Rebecca Jackson, Jeff Klein, Kenda Knowles, Brian Koopman, Arthur Kosowsky, Vincent Lakey, Dale Li, Yaqiong Li, Zack Li, Martine Lokken, Thibaut Louis, Marius Lungu, Amanda MacInnis, Mathew Madhavacheril, Felipe Maldonado, Maya Mallaby-Kay, Danica Marsden, Jeff McMahon, Felipe Menanteau, Kavilan Moodley, Tim Morton, Toshiya Namikawa, Federico Nati, Laura Newburgh, John P. Nibarger, Andrina Nicola, Michael D. Niemack, Michael R. Nolta, John Orlowski-Sherer, Lyman A. Page, Christine G. Pappas, Bruce Partridge, Phumlani Phakathi, Giampaolo Pisano, Heather Prince, Roberto Puddu, Frank J. Qu, Jesus Rivera, Naomi Robertson, Felipe Rojas, Maria Salatino, Emmanuel Schaan, Alessandro Schillaci, Neelima Sehgal, Blake D. Sherwin, Carlos Sierra, Jon Sievers, Cristobal Sifon, Precious Sikhosana, Sara Simon, David N. Spergel, Suzanne T. Staggs, Jason Stevens, Emilie Storer, Dhaneshwar D. Sunder, Eric R. Switzer, Ben Thorne, Robert Thornton, Hy Trac, Jesse Treu, Carole Tucker, Leila R. Vale, Alexander Van Engelen, Jeff Van Lanen, Eve M. Vavagiakis, Kasey Wagoner, Yuhan Wang, Jonathan T. Ward, Edward J. Wollack, Zhilei Xu, Fernando Zago, and Ningfeng Zhu. The Atacama Cosmology Telescope: DR4 maps and cosmological parameters. , 2020(12):047, December 2020. doi: 10.1088/1475-7516/2020/12/047.

[151] T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, S. Allen, J. Annis, S. Avila, D. Bacon, K. Bechtol, A. Bermeo, G. M. Bernstein, E. Bertin, S. Bhargava, S. Bocquet, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, X. Chen, A. Choi, M. Costanzi, M. Crocce, L. N. da Costa, T. M. Davis, J. De Vicente, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, A. Drlica-Wagner, K. Eckert, T. F. Eifler, J. Elvin-Poole, J. Estrada, S. Everett, A. E. Evrard, A. Farahi, I. Ferrero, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, P. Giles, S. Grandis, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. Huterer, D. J. James, M. Jarvis, T. Jeltema, M. W. G. Johnson, M. D. Johnson, S. Kent, E. Krause, R. Kron, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G.

Maia, A. Mantz, J. L. Marshall, P. Martini, J. Mayers, P. Melchior, J. Mena-Fernández, F. Menanteau, R. Miquel, J. J. Mohr, R. C. Nichol, B. Nord, R. L. C. Ogando, A. Palmese, F. Paz-Chinchón, A. A. Plazas, J. Prat, M. M. Rau, A. K. Romer, A. Roodman, P. Rooney, E. Rozo, E. S. Rykoff, M. Sako, S. Samuroff, C. Sánchez, E. Sanchez, A. Saro, V. Scarpine, M. Schubnell, D. Scolnic, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, J. Allyn. Smith, M. Smith, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, M. A. Troxel, D. L. Tucker, T. N. Varga, A. von der Linden, A. R. Walker, R. H. Wechsler, J. Weller, R. D. Wilkinson, H. Wu, B. Yanny, Y. Zhang, Z. Zhang, J. Zuntz, and DES Collaboration. Dark Energy Survey Year 1 Results: Cosmological constraints from cluster abundances and weak lensing. , 102(2):023509, July 2020. doi: 10.1103/PhysRevD.102.023509.

[152] F. Pacaud, M. Pierre, J. B. Melin, C. Adami, A. E. Evrard, S. Galli, F. Gastaldello, B. J. Maughan, M. Sereno, S. Alis, B. Altieri, M. Birkinshaw, L. Chiappetti, L. Faccioli, P. A. Giles, C. Horellou, A. Iovino, E. Koulouridis, J. P. Le Fèvre, C. Lidman, M. Lieu, S. Maurogordato, L. Moscardini, M. Plionis, B. M. Poggianti, E. Pompei, T. Sadibekova, I. Valtchanov, and J. P. Willis. The XXL Survey. XXV. Cosmological analysis of the C1 cluster number counts. , 620:A10, November 2018. doi: 10.1051/0004-6361/201834022.

[153] Catherine Heymans, Tilman Tröster, Marika Asgari, Chris Blake, Hendrik Hildebrandt, Benjamin Joachimi, Konrad Kuijken, Chieh-An Lin, Ariel G. Sánchez, Jan Luca van den Busch, Angus H. Wright, Alexandra Amon, Maciej Bilicki, Jelte de Jong, Martin Crocce, Andrej Dvornik, Thomas Erben, Maria Cristina Fortuna, Fedor Getman, Benjamin Giblin, Karl Glazebrook, Henk Hoekstra, Shahab Joudaki, Arun Kannawadi, Fabian Köhlinger, Chris Lidman, Lance Miller, Nicola R. Napolitano, David Parkinson, Peter Schneider, HuanYuan Shan, Edwin A. Valentijn, Gijs Verdoes Kleijn, and Christian Wolf. KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. , 646:A140, February 2021. doi: 10.1051/0004-6361/202039063.

[154] James W Beletic, Richard Blank, David Gulbransen, Donald Lee, Markus Loose, Eric C Piquette, Thomas Sprafke, William E Tennant, Majid Zandian, and Joseph Zino. Teledyne imaging sensors: infrared imaging technologies for astronomy and civil space. In *High Energy, Optical, and Infrared Detectors for Astronomy III*, volume 7021, page 70210H. International Society for Optics and Photonics, 2008.

[155] G L Hansen, JL Schmit, and TN Casselman. Energy gap versus alloy composition and temperature in hg1- x cd x te. *Journal of Applied Physics*, 53(10):7099–7101, 1982.

[156] JC Clémens, B Serra, M Niclas, A Ealet, W Gillard, A Secroun, R Barbier, B Kubik, S Ferriol, G Smadja, et al. Euclid detector system demonstrator model: a first demonstration

of the nisp detection system. In *UV/Optical/IR Space Telescopes and Instruments: Innovative Technologies and Concepts VII*, volume 9602, page 96020Y. International Society for Optics and Photonics, 2015.

[157] L Dressel. Wide field camera 3 instrument handbook for cycle 21 v. 5.0. *Wide Field Camera 3*, 2012.

[158] Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, 2015.

[159] David J Bacon, Alexandre R Refregier, and Richard S Ellis. Detection of weak gravitational lensing by large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 318 (2):625–640, 2000.

[160] Nick Kaiser, Gillian Wilson, and Gerard A Luppino. Large-scale cosmic shear measurements. *arXiv preprint astro-ph/0003338*, 2000.

[161] Ludovic Van Waerbeke, Y Mellier, T Erben, JC Cuillandre, F Bernardeau, R Maoli, E Bertin, HJ Mc Cracken, O Le Fevre, B Fort, et al. Detection of correlated galaxy ellipticities on cfht data: first evidence for gravitational lensing by large-scale structures. *Arxiv preprint astro-ph/0002500*, 2000.

[162] David M Wittman, J Anthony Tyson, David Kirkman, Ian Dell'Antonio, and Gary Bernstein. Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. *nature*, 405(6783):143–148, 2000.

[163] N Mauri, S Dusini, F Fornari, D Di Ferdinando, F Giacomini, F Laudisio, L Patrizii, C Sirignano, G Sirri, L Stanco, et al. The euclid near infrared spectro-photometer (nisp) instrument and science. In *Journal of Physics: Conference Series*, volume 1342, page 012122. IOP Publishing, 2020.

[164] Daniel J Eisenstein, Idit Zehavi, David W Hogg, Roman Scoccimarro, Michael R Blanton, Robert C Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, et al. Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *The Astrophysical Journal*, 633(2):560, 2005.

[165] Nabila Aghanim, Yashar Akrami, Frederico Arroja, Mark Ashdown, J Aumont, Carlo Baccigalupi, M Ballardini, Anthony J Banday, RB Barreiro, Nicola Bartolo, et al. Planck 2018 results-i. overview and the cosmological legacy of planck. *Astronomy & Astrophysics*, 641:A1, 2020.

[166] Laura Salvati, Marian Douspis, and Nabila Aghanim. Impact of systematics on cosmological parameters from future galaxy cluster surveys. , 643:A20, November 2020. doi: 10.1051/0004-6361/202038465.

[167] AM Fowler and Ian Gatley. Demonstration of an algorithm for read-noise reduction in infrared arrays. *The Astrophysical Journal*, 353:L33, 1990.

[168] Bogna Kubik, Rémi Barbier, Alain Castera, Eric Chabanat, Sylvain Ferriol, and Gérard Smadja. Optimization of the multiple sampling and signal extraction in nondestructive exposures. *Journal of Astronomical Telescopes, Instruments, and Systems*, 1(3):038001, 2015.

[169] G. Smadja, C. Cerna, A. Castera, and A. Ealet. Frequency analysis of the noise in the fowler(n) sampling of a h2rg(2k ×2k) near-ir detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 622(1):288 – 294, 2010. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2010.06.271. URL http://www.sciencedirect.com/science/article/pii/S0168900210014610.

[170] Bogna Kubik, Remi Barbier, Alain Castera, Eric Chabanat, Sylvain Ferriol, and Gerard Smadja. Predictive model of the temporal noise correlations in hgcdte array. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 787:315 – 318, 2015. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2015.01.011. URL http://www.sciencedirect.com/science/article/pii/S0168900215000376. New Developments in Photodetection NDIP14.

[171] Rachel E Anderson and Karl D Gordon. Optimal cosmic-ray detection for nondestructive read ramps. *Publications of the Astronomical Society of the Pacific*, 123(908):1237, 2011.

[172] Konstantin Protassov. *Analyse statistique de données expérimentales*. EDP Sciences, 2012.

[173] S. H. Moseley, Richard G. Arendt, D. J. Fixsen, Don Lindler, Markus Loose, and Bernard J. Rauscher. Reducing the read noise of H2RG detector arrays: eliminating correlated noise with efficient use of reference signals. In , volume 7742 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 77421B, Jul 2010. doi: 10.1117/12.866773.

[174] Bernard J. Rauscher, S. H. Moseley, R. G. Arendt, D. Fixsen, D. Lindler, and M. Loose. Reducing the Read Noise of the James Webb Near Infrared Spectrograph by Improved Reference Sampling & Subtraction (IRS-square). In *American Astronomical Society Meeting Abstracts #219*, volume 219 of *American Astronomical Society Meeting Abstracts*, page 328.06, Jan 2012.

[175] Matthew Newville and Till Stensitzki. Non-linear least-squares minimization and curve-fitting for python matthew newville. *Till Stensitzki, and others*, 2017.

[176] Bogna Kubik, Remi Barbier, Eric Chabanat, Arnaud Chapon, Jean-Claude Clemens, Anne Ealet, Sylvain Ferriol, William Gillard, Aurelia Secroun, Benoit Serra, Gerard Smadja, and André Tilquin. A New Signal Estimator from the NIR Detectors of the Euclid Mission. , 128(968):104504, October 2016. doi: 10.1088/1538-3873/128/968/104504.

[177] Alexander I Merson, Carlton M Baugh, John C Helly, Violeta Gonzalez-Perez, Shaun Cole, Richard Bielby, Peder Norberg, Carlos S Frenk, Andrew J Benson, Richard G Bower, et al. Lightcone mock catalogues from semi-analytic models of galaxy formation–i. construction and application to the bzk colour selection. *Monthly Notices of the Royal Astronomical Society*, 429(1):556–578, 2013.

[178] Shaun Cole, Cedric G Lacey, Carlton M Baugh, and Carlos S Frenk. Hierarchical galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 319(1):168–204, 2000.

[179] RG Bower, AJ Benson, R Malbon, JC Helly, CS Frenk, CM Baugh, Shaun Cole, and Cedric G Lacey. Breaking the hierarchy of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 370(2):645–655, 2006.

[180] Lilian Jiang, John C Helly, Shaun Cole, and Carlos S Frenk. N-body dark matter haloes with simple hierarchical histories. *Monthly Notices of the Royal Astronomical Society*, 440 (3):2115–2135, 2014.

[181] Michel Fioc and Brigitte Rocca-Volmerange. Pegase: a uv to nir spectral evolution model of galaxies-application to the calibration of bright galaxy counts. *arXiv preprint astro-ph/9707017*, 1997.

[182] Gary A Mamon, Andrea Biviano, and Giuseppe Murante. The universal distribution of halo interlopers in projected phase space-bias in galaxy cluster concentration and velocity anisotropy? *Astronomy & Astrophysics*, 520:A30, 2010.

[183] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, January 2010. doi: 10.2140/camcos.2010.5.65.

[184] Daniel Foreman-Mackey, Alex Conley, Will Meierjurgen Farr, David W. Hogg, Dustin Lang, Phil Marshall, Adrian Price-Whelan, Jeremy Sanders, and Joe Zuntz. emcee: The MCMC Hammer, March 2013.

[185] M Ricci, C Benoist, S Maurogordato, C Adami, LUCIO Chiappetti, FABIO Gastaldello, V Guglielmo, B Poggianti, Mauro Sereno, R Adam, et al. The xxl survey-xxviii. galaxy luminosity functions of the xxl-n clusters. *Astronomy & Astrophysics*, 620:A13, 2018.

[186] M Pierre, F Pacaud, C Adami, S Alis, B Altieri, N Baran, C Benoist, M Birkinshaw, ANGELA Bongiorno, MN Bremer, et al. The xxl survey-i. scientific motivations- xmm-newton observing plan- follow-up observations and simulation programme. *Astronomy & Astrophysics*, 592:A1, 2016.

[187] C Chang, M Jarvis, B Jain, SM Kahn, D Kirkby, A Connolly, S Krughoff, E-H Peng, and JR Peterson. The effective number density of galaxies for weak lensing measurements in the lsst project. *Monthly Notices of the Royal Astronomical Society*, 434(3):2121–2135, 2013.

[188] A Gonzalez. Cluster detection via wavelets. *Building the Euclid Cluster Survey-Scientific Program*, page 7, 2014.

[189] Carlo Giocoli, Eric Jullo, R Benton Metcalf, Sylvain De La Torre, Gustavo Yepes, Francisco Prada, Johan Comparat, Stefan Göttlober, Anatoly Kyplin, Jean-Paul Kneib, et al. Multidarklens simulations: weak lensing light-cones and data base presentation. *Monthly Notices of the Royal Astronomical Society*, 461(1):209–223, 2016.

[190] Peter S Behroozi, Risa H Wechsler, and Hao-Yi Wu. The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores. *The Astrophysical Journal*, 762(2):109, 2012.

[191] Alexander M Beck, Giuseppe Murante, Alexander Arth, R-S Remus, Adelheid F Teklu, Julius MF Donnert, Susana Planelles, Marcus C Beck, Pascal Förster, Maximilian Imgrund, et al. An improved sph scheme for cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 455(2):2110–2130, 2016.

[192] Francesco Haardt and Piero Madau. Radiative transfer in a clumpy universe: Ii. the utraviolet extragalactic background. *arXiv preprint astro-ph/9509093*, 1995.

[193] Robert PC Wiersma, Joop Schaye, and Britton D Smith. The effect of photoionization on the cooling rates of enriched, astrophysical plasmas. *Monthly Notices of the Royal Astronomical Society*, 393(1):99–107, 2009.

[194] Luca Tornatore, Stefano Borgani, K Dolag, and F Matteucci. Chemical enrichment of galaxy clusters from hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 382(3):1050–1072, 2007.

[195] Gilles Chabrier. Galactic stellar and substellar initial mass function. *Publications of the Astronomical Society of the Pacific*, 115(809):763, 2003.

[196] Volker Springel and Lars Hernquist. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. *Monthly Notices of the Royal Astronomical Society*, 339(2):289–311, 2003.

[197] Lisa K Steinborn, Klaus Dolag, Michaela Hirschmann, M Almudena Prieto, and Rhea-Silvia Remus. A refined sub-grid model for black hole accretion and agn feedback in large cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 448 (2):1504–1525, 2015.

[198] Steffen R Knollmann and Alexander Knebe. Ahf: Amiga's halo finder. *The Astrophysical Journal Supplement Series*, 182(2):608, 2009.

[199] Julien EG Devriendt, Bruno Guiderdoni, and Rachida Sadat. Galaxy modelling–i. spectral energy distributions from far-uv to sub-mm wavelengths. *arXiv preprint astro-ph/9906332*, 1999.

[200] K Dolag, Stefano Borgani, G Murante, and V Springel. Substructures in hydrodynamical cluster simulations. *Monthly Notices of the Royal Astronomical Society*, 399(2):497–514, 2009.

[201] A. Jiménez Muñoz, J. Macías-Pérez, A. Secroun, W. Gillard, B. Kubik, N. Auricchio, A. Balestra, C. Bodendorf, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, V. Capobianco, C. Carbone, J. Carretero, R. Casas, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, L. Conversi, Y. Copin, L. Corcione, A. Costille, M. Cropper, H. Degaudenzi, M. Douspis, F. Dubath, S. Dusini, A. Ealet, E. Franceschi, P. Franzetti, M. Fumana, B. Garilli, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, F. Hormuth, K. Jahnke, S. Kermiche, A. Kiessling, M. Kilbinger, M. Kümmel, M. Kunz, H. Kurki-Suonio, R. Laureijs, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, R. Massey, E. Medinaceli, S. Mei, M. Meneghetti, G. Meylan, L. Moscardini, S. M. Niemi, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, W. J. Percival, S. Pires, G. Polenta, M. Poncet, L. Popa, L. Pozzetti, F. Raison, R. Rebolo, M. Roncarelli, E. Rossetti, R. Saglia, M. Sauvage, R. Scaramella, P. Schneider, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, D. Tavagnacco, A. N. Taylor, H. I. Teplitz, I. Tereno, R. Toledo-Moreo, L. Valenziano, T. Vassallo, G. A. Verdoes Kleijn, Y. Wang, J. Weller, M. Wetzstein, G. Zamorani, and J. Zoubian. Euclid: Estimation of the Impact of Correlated Readout Noise for Flux Measurements with the Euclid NISP Instrument. , 133 (1027):094502, September 2021. doi: 10.1088/1538-3873/ac21de.

[202] ES Rykoff, Eduardo Rozo, MT Busha, CE Cunha, A Finoguenov, A Evrard, J Hao, BP Koester, A Leauthaud, B Nord, et al. redmapper. i. algorithm and sdss dr8 catalog. *The Astrophysical Journal*, 785(2):104, 2014.

[203] Ting-Wen Lan, Brice Ménard, and Houjun Mo. The galaxy luminosity function in groups and clusters: the faint-end upturn and the connection to the field luminosity function. , 459(4):3998–4019, July 2016. doi: 10.1093/mnras/stw898.

## Résumé

*Euclid* est une mission satellite de classe moyenne qui sera lancée par l'ESA en 2023 et qui est composée de deux instruments : l'instrument Visible (VIS) et le spectromètre et photomètre dans le proche infrarouge (NISP). Il est principalement consacré à la cosmologie et entend dévoiler la nature de l'énergie et de la matière noire en utilisant diverses sondes cosmologiques, comme par exemple la distribution des amas de galaxies en masse et en redshift. Cette thèse se concentre d'abord sur l'estimation des implications du bruit de lecture corrélé dans les détecteurs NISP pour les mesures finales de flux en vol. Nous trouvons que pour le fond de ciel attendu, l'algorithme actuel à bord du satellite fournira des estimations de flux non biaisées. Deuxièmement, sur la construction d'un catalogue synthétique des amas des galaxies par des modèles analytiques qui reproduisent les propriétés observationnelles des galaxies d'amas telles que: la fonction de luminosité et leur distribution de densité. Ces propriétés sont déduites de deux catalogues simulés: un catalogue mock semi-analytique (SAMs) fourni par la Collaboration *Euclid*, et un catalogue d'amas simulé hydrodynamique, provenant du projet 300th. Enfin, notre catalogue synthétique est utilisé pour déduire la fonction de sélection, qui est essentielle pour calculer le nombre d'amas, en utilisant une technique d'injection d'amas. Nous concluons que la technique d'injection d'amas est une alternative prometteuse aux méthodes de simulation de remplissage. En outre, nous constatons que dans les deux cas, une attention particulière est nécessaire pour reproduire les amas et les propriétés des galaxies d'arrière-plan.



## Abstract

*Euclid* is a Medium Class satellite mission to be launched by ESA in 2023 and composed of two instruments: the Visible instrument (VIS) and the Near Infrared Spectrometer and Photometer (NISP). It is mainly devoted to cosmology and intends to unveil the nature of the Dark Energy and the Dark Matter by using various cosmological probes, such as for example the galaxy clusters distribution as a function of their mass and redshift. This thesis is focused, first, in the estimation of the implications of correlated readout noise in the NISP detectors for the final in-flight flux measurements. We find that for the expected sky background the current on board algorithm will provide unbiased flux estimates. Second, in the construction of a synthetic catalogue of galaxy clusters using analytical models that reproduce cluster galaxy observational properties such as: the luminosity function and their galaxy density distribution. These properties are inferred from two simulated catalogues: a semi-analytical (SAMs) mock catalogue as provided by *Euclid* Collaboration, and a hydrodynamical simulated cluster catalogue, coming from the 300th Project. Finally, our synthetic catalogue is used for inferring the selection function, which is essential for computing the cluster number counts, using a cluster injection technique. We conclude that cluster injection technique is a promising alternative to fill simulation methods. Furthermore, we find that in both cases special care is needed to reproduce the clusters and background properties.