Pablo Andrés Araya Araya

Protocluster Detection in Mock Photometric Surveys

São Paulo

2020

Pablo Andrés Araya Araya

Protocluster Detection in Mock Photometric Surveys

Versão Corrigida. O original encontra-se disponível na Unidade

Dissertação apresentada ao Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Área de concentração: Astronomia extragalática.

Orientador: Prof. Dr. Laerte Sodré Júnior.

São Paulo
2020

Tanto fervor tiene el cielo,
tanto ama, tanto regala,
que a veces yo quiero más
la noche que las mañanas.

Gabriela Mistral - Noche Andina

Agradecimentos

I think that this is the hardest part of the MSc Dissertation to write because I have many people to acknowledge. Many people supported me in the academic context, social life, and economically. I will divide this section into three languages to make it easier for people that do not speak English, Portuguese, or Spanish.

Additionally, I have to thank Roderik Overzier, who always was there, open for any type of question. Roderik, you helped us a lot to improve our works, and thank you for inviting us to all those amazing events.

To finish the English part, I need to mention my sister Julia (she does not speak Spanish). We had the possibility to met each other just at the end of last year when I was at the peak of stress. Meet you, after more than 10 years since I know that I have a sister, gave me a lot of force to continue. Hope to see you soon!

Agora em Português. Muitas pessoas foram muito legais comigo desde o começo. Se fosse por mim, eu agradeceria toda pessoa que teve contato comigo, incluindo pessoal da segurança, limpeza, administrativos, etc. mas não posso escrever 10 folhas de agradecimentos.

A primeira pessoa é obviamente o meu orientador, professor Laerte Sodré Jr., principalmente pela paciência, porque eu sei que sou meio chato as vezes. Obrigado professor por tudo, aprendi mais do que eu imaginei durante esse processo trabalhando com você. Também, muito obrigado pela sua ajuda econômica quando eu mais precisei. Durante o meu mestrado eu não trabalhei sozinho, trabalhei diretamente com o Marcelo, que virou um parceiro não só de pesquisa, mas também da vida (tamo junto!). Aproveito para agradecer o nosso grupo tudo de pesquisa (estudantes do Laerte) pelas boas discussões nas reuniões. Também ao Erik pela sua ajuda com os photo-zs, amizade, e os coffee-breaks. Gostaria de agradecer também aos professores das matérias que eu fiz, em particular ao prof. Eduardo Cypriano com quem além de aprender bastante nas aulas dele, jogamos tênis.

Quando cheguei em São Paulo, o Johnny, que quase nem conhecia, hospedou-me por quase 10 meses no apartamento dele. Obrigado cara pela sua amabilidade e aquelas boas conversas. O Fabito também abriu as portas da casa dele para mim quando precisei. Mas não só isso, acho que você é o meu melhor amigo no Brasil junto com a Thayse. Estou devendo um ceviche para vocês. Bom, não posso deixar de agradecer a Stephane (sim, a minha ex) e a família dela por me apoiar em praticamente tudo durante quase os meus primeiros 2 anos no Brasil.

Como não agradecer ao pessoal da melhor sala do IAG, a sala B-304 (sala da Mirian), obrigado Mirian, Henrique, Raniere e André por fazer da nossa sala o lugar perfeito do mundo para trabalhar.

De longe, uma das melhores coisas que aconteceram comigo nesse ano foi virar mais amigo dos meu amigos. Vocês foram meu soporte quando achei que ia colapsar. Obrigado aos já mecionados mais o Ge, Natalia, Maria Luisa, Lilianne, Jhon, Roberta, Julia, Fer, Amanda e todos os que compartilharam comigo pelo menos uma cerveja. Vocês não tem noção das saudades que tenho das nossas visitas ao Beco da USP, os rolês infinitos na casa da Thayse, aqueles no apartamento do Fabito, coffee-breaks na sala do Erik, o bandejão, tudo! Obrigado à todos vocês por fazer o fato de sair da minha cidade uma experiência maravilhosa!

Terminemos en español, más correcto en chileno. Salir de mi querido Coquimbo para llegar a São Paulo y tambien desde la Universidad de La Serena (ULS) a la Universidade de São Paulo (USP) fue un cambio muy cuático. Aunque la ULS está lejos de ser de las mejores instituciones de Chile y la USP ser la mejor de Latinoamerica, nunca sentí que me faltó algo por aprender en el pregrado. Llevar a cabo de manera satisfactoria (para mi) este magíster se lo debo también a mis profesores de la ULS, principalmente al Nilo y profe Héctor. Con ellos fue con los que más aprendí a hacer investigación, programar y pensar de manera crítica.

Echo de menos todo y a todos de ULS, empezando por el ping-pong diario a la hora de almuerzo con don Jaime, don Luis, Pacheco y Fabio. Obvio que a mis amigos que me apoyaron bastante durante todo el proceso del magister: Daniel, Nahir (me hubiera gustado estar para el matrimonio), Alex, Pallero (papi que pasó?), Panchito, los Gs, Nico, Sebi (aguante Coquimbo), Andrea, Pablo, Elias, Maripan, Yilin, Arlin, y muchos más, por esos carretes y converzas que se extrañan tanto, los quiero caleta.

Encuentro muy bacán aún tener amigos del colegio. Si bien nos distanciamos físicamente, cuando nos juntamos es como si nada hubiera cambiado. Gracias Raúl, Vanni, Manolito, Vero y Dante por apoyarme siempre.

Quiero agradecer tambien a toda mi familia. De parte de mi papá, agradezco principalmente, bueno, a él, a mi tía Teresa y a mi abuela por motivarme constantemente desde que los conocí. Ahora los que siempre han estado conmigo, a mis dos abuelos (el tata y la mama), mis primitos (Elisa, Emilio, Consuelo y Gabriel), mis tíos (tía Carla, tío Andrés, Sergio y Johana). No saben lo difícil que es pasar de verlos todos los domingos a sólo unos días al año, gracias por todo. Para finalizar, quiero agradecer a mi mamá, la persona que siempre está ahi, que siempre me apoya y me entiende. Te dedico esta tesis a ti!

Resumo

Os progenitores dos aglomerados de galáxias atuais fornecem pistas importantes sobre a evolução das estruturas em grande escala, o crescimento de suas massas e a evolução de galáxias. Simulações são uma ferramenta importante para esses estudos, pois são usadas para interpretar observações de proto-aglomerados que são detectados como excessos de densidade na distribuição de galáxias e além disso, testar cenários evolutivos.

Nesta dissertação de mestrado, apresentamos um conjunto de "protocluster-lightcones", apelidados de PCcones. Eles são catálogos de galáxias simulados gerados a partir da Simulação Millennium com a versão anterior do modelo semi-analítico `L-GALAXIES`. Esses cones de luz foram construídos colocando uma estrutura desejada no redshift de interesse ao centro do cone e levando em consideração os vínculos observacionais associados a um determinado levantamento. Verificamos que os redshifts fotométricos (foto-zs) produzidos com o PCcones são melhores que os obtidos diretamente da simulação Millennium. Determinamos a precisão esperada na detecção de proto-aglomerados usando foto-zs na faixa de $z = 1 - 3$ no HSC-SSP e na previsão de 10 anos do LSST. Com nossa técnica, esperamos recuperar $\sim 38\%$ e 42% de todos os progenitores de aglomerados de galáxias massivos com mais de 70% de pureza para HSC-SSP e LSST, respectivamente, no intervalo de redshift $z = 1 - 3$.

Nossos resultados também indicam que a combinação de vínculos observacionais e incertezas nos foto-z afetam criticamente a detecção de estruturas. Isso acontece para ambas as amostras emuladas.

Também comparamos nossas emulações (mocks) do Deep CFHTLS em $z \leqslant 1.5$ com catálogos de aglomerados observados, como uma validação extra dos mocks e de nossos métodos. Descobrimos que as duas distribuições são consistentes uma com a outra. Isso indica que com os PCcones, podemos reproduzir resultados observacionais satisfatoriamente.

Palavras-chave: métodos: numericos – galáxias: aglomerados: geral – galáxias: alto-redshift

Abstract

The progenitors of present-day galaxy clusters give important clues about the evolution of the large scale structure, cosmic mass assembly, and galaxy evolution. Simulations are a major tool for these studies since they are used to interpret observations of protoclusters that are detected as enhancements in the distribution of galaxies, and test evolutionary scenarios.

In this MSc Dissertation, we introduce a set of "protocluster-lightcones", dubbed PCcones. They are mock galaxy catalogs generated from the Millennium Simulation with the previous version of the L-GALAXIES semi-analytic model. These lightcones were constructed by placing a desired structure at the redshift of interest in the center of the cone and taking into account the observational constraints associated with a given survey. We show that photometric redshifts (photo-zs) obtained with PCcones are more accurate than those obtained directly with the Millennium Simulation. We determine the expected accuracy of protocluster detection using photo-zs in the $z = 1 - 3$ range in the wide layer of HSC-SSP and the 10-year LSST forecast. With our technique, we expect to recover $\sim 38\%$ and 43% of all massive galaxy cluster progenitors with more than 70% of purity for HSC-SSP and LSST, respectively, at the $z = 1 - 3$ redshift interval.

Our results also indicate that the combination of observational constraints and photo-z uncertainties affects the detection of structures critically. This happens for both emulated samples.

We also compare our mocks of the Deep CFHTLS at $z \leqslant 1.5$ with observed cluster catalogs, as an extra validation of the lightcones and our methods. Here, we found that both distributions are consistent with each other. This indicates that with PCcones, we can reproduce satisfactorily observational results.

Key words: methods: numerical – galaxies: clusters: general – galaxies: high-redshift

Contents

List of Figures

List of Figures

List of Figures

List of Tables

1   Introduction

Protoclusters are the progenitors of the present-day most massive virialized (or quasi) structures. In our current model of structure formation, the ΛCDM framework, they are an ensemble of dark matter halos at high redshifts that will merge, forming galaxy clusters. Their environment and properties make them exciting targets to study galaxy formation and evolution, cosmology, and the mass assembly in the universe at early times. Current and future photometric surveys could increase the number of protocluster candidates, which is still limited. In this chapter, we shall discuss some important aspects such as the ΛCDM cosmological model and its observational support, which are the base of simulations. We will also explain the main features of the structure evolution models, as well as how they explain the large scale structures of the universe. Finally, we will characterize clusters but focusing on protoclusters, since these are the objects that we want to investigate.

1.1   │   Observational Support of ΛCDM

[1]The present-day most acceptable cosmological model is the ΛCDM (Λ Cold Dark Matter). It assumes that the universe is in an accelerated expansion produced by a dark energy in the form of a cosmological constant, Λ, and that most of the matter is in the form of dark matter, elementary particles with non-relativistic velocities. For these reasons, this component is called Cold Dark Matter (CDM). Additionally, the inflation model proposes that $\sim 10^{-35}$ seconds after the Big Bang the universe suffered an exponential expansion driven by a kind of vacuum energy. The cosmic inflation generates quantum fluctuations, which gravity will amplify, triggering the formation of galaxies and the large scale structure of the universe.

The current paradigm assumes a flat ΛCMD universe with $\Omega_m \sim 0.3$, $\Omega_\Lambda \sim 0.7$ and $h \sim 0.7$, where Ω_m and Ω_Λ are the density of mass and dark energy in units of a critical density and h the Hubble constant in units of 100 km/s/Mpc (Spergel et al., 2003). In this section, we describe three observational results that support the ΛCDM cosmological model, starting with the nature of the dark matter, following with the discovery of the dark energy, and finalizing with the flatness of the universe.

---

[1]Most of the content that will be described in this section and the next one comes from Schneider (2006).

1.1.1 | The Power Spectrum and the Nature of Dark Matter

One of the main tests of the cosmological model is the power spectrum, which describes the density contrast of the universe in different scales. Different models predict distinct forms for this distribution, that depends if the dark matter is "cold" or "hot". We can define the power spectrum, $P(k)$, as the Fourier transform of the correlation function, $\xi(r)$:

$$P(k) = 2\pi \int_0^\infty r^2 \frac{\sin kr}{kr} \xi(r) dr \tag{1.1}$$

where r indicates the separation between two galaxies and k is the wave number (which is proportional to the inverse of the length-scale). The correlation function (or two-point correlation function), $\xi(r)$, is the excess of probability of finding a galaxy separated by the distance r from another one, with respect to we expect for randomly distributed galaxies in a given volume.
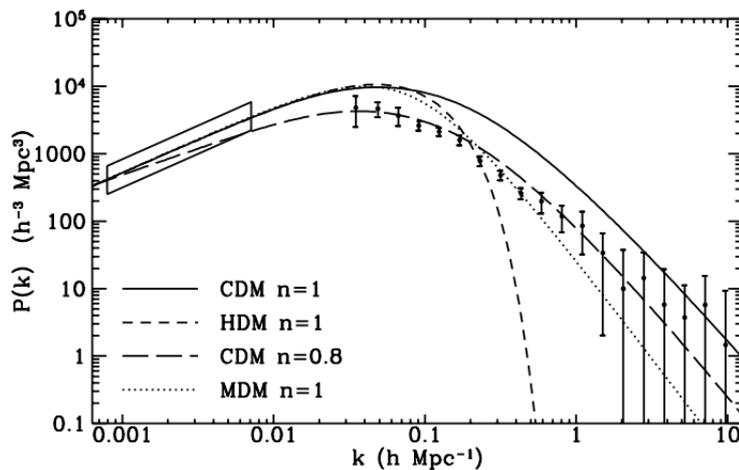
Figure 1.1: Predicted matter power spectrum, assuming a hot (small-dashed line), warm (dotted), and cold (solid and long-dashed) dark matter. n is the slope of the primordial power spectrum ($P(k) = Ak^n$). Dots with error bars represent observational data. Figure from Kolb (1998).

Since the density in the universe evolves with time, the power spectrum also has a temporal dependence. Therefore, the model power spectrum depends on the cosmological model as well as on the theory of structure evolution. Another determining factor to consider is the nature of the dark matter. We can distinguish between three main types of dark matter: hot, warm, and cold dark matter (HDM, WDM, and CDM, respectively), and they differ in their velocities/"temperature". Different types of dark matter imply in different particle masses; for example, the particle mass in the HDM model needs to be light, to achieve relativistic velocities. Accordingly to Schneider (2006), the dark matter particle assuming a HDM model has to be of order $mc^2 \sim 10 - 100$ eV. We expect the opposite in the CDM universe, with a particle mass of $mc^2 \sim 10 - 1000$ GeV (Bertone and Tait, 2018, assuming a weakly interacting massive particle scenario), that will never attain relativistic velocities.

Figure 1.1 shows the theoretical power spectrum for the three different dark matter models

above, compared with a compilation of several observational results. The disagreement of the HDM power spectrum with the observations led to the rejection of the hypothesis that the dark matter particles present velocities $v \sim c$. Besides, the HDM model suggests a different scenario of structure evolution. If the dark matter consists of relativistic particles, these would not be gravitationally bound by the potential well of dense regions in the universe. Therefore, the dark matter particles are free to escape from overdensities, generating a fragmentation of large scale systems; we call this effect *free-streaming*. It implies that small scale structures in the HDM model come from the dissolution of the largest ones. Contrarily, the CDM model suggests that minor structures were formed first and the largest ones by the merger of those (hierarchical model).

1.1.2 | Supernovae type Ia and the Dark Energy

We link the cosmological constant, Λ, with the acceleration of the expansion of the universe, as a form of dark energy. Here, we will describe the observational results that drove the discovery of dark energy.

Type Ia supernovae are triggered in binary systems when a white dwarf star accretes a specific quantity of mass from its companion (the limit of Chandrasekhar). We suppose that this limit is the same for all white dwarfs. Therefore, we also expect that the maximum luminosity produced by supernovae type Ia is the same. Then, if we know the redshift of the host galaxy, we can constrain the cosmological parameters, since the distance modulus depends on the luminosity distance as follow:

$$\mu(z) \equiv m - M = 5\log(d_L(z)/10pc) + K(z) + A \tag{1.2}$$

where m and M are the apparent and absolute magnitudes, respectively; $K(z)$ is the K-correction and A the reddening. $d_L(z)$ is the luminosity distance defined as: $d_L(z) = (1+z)d_M(z)$, where $d_M(z)$ is different depending on the curvature of the universe:

$$d_M(z) = \left\{ \begin{array}{ll} \frac{d_H}{\sqrt{\Omega_k}}\sinh\left(\sqrt{\Omega_k}d_C(z)/d_H\right) & ;\text{for } \Omega_k > 0 \\ d_C(z) & ;\text{for } \Omega_k = 0 \\ \frac{d_H}{\sqrt{|\Omega_k|}}\sin\left(\sqrt{|\Omega_k|}d_C(z)/d_H\right) & ;\text{for } \Omega_k < 0 \end{array} \right\} \tag{1.3}$$

with:

$$d_C(z) = \frac{c}{H_0}\int_0^z \frac{dz'}{E(z')} = d_H \int_0^z \frac{dz'}{E(z')} \tag{1.4}$$

where $E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}$, with $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$.

The maximum absolute magnitude derived from the SN Ia light curves is not explicitly a constant, and the difference could reach ~ 1.5 mag. However, Phillips (1993) proposed empirical color corrections based on the width of the observed light curves. Hamuy et al. (1996) confirmed these proposed corrections using a set of SN Ia data, where the dispersion of the peak absolute

magnitudes decrease to ~ 0.15 mag. Riess et al. (1998) and Perlmutter et al. (1999) (two different groups) used type Ia supernovae as standard candles to constrain the cosmological parameters. Both obtained that high redshift SNe Ia are fainter than the values predicted, assuming a standard Friedmann-Lemaître cosmology. This suggests that the universe is in accelerated expansion, since the most probable values for Ω_Λ and Ω_m are 0.76 and 0.24, assuming a flat universe.

We reproduce this experiment by fitting 580 already corrected (by reddening and K-corrections) SNe Ia distance modulus to Equation 1.2, as one of the activities for the Extragalactic Astronomy course. We used the data presented in Suzuki et al. (2012) [2] We performed a MCMC algorithm with 10,000 iterations. Figure 1.2 shows the probability density distribution in the Ω_m and Ω_Λ space, considering the last 7,000 values. Finally, we have obtained $(\Omega_m, \Omega_\Lambda) = (0.28 \pm 0.02, 0.72 \pm 0.03)$. These values are similar to those obtained by Riess et al. (1998) and Perlmutter et al. (1999).
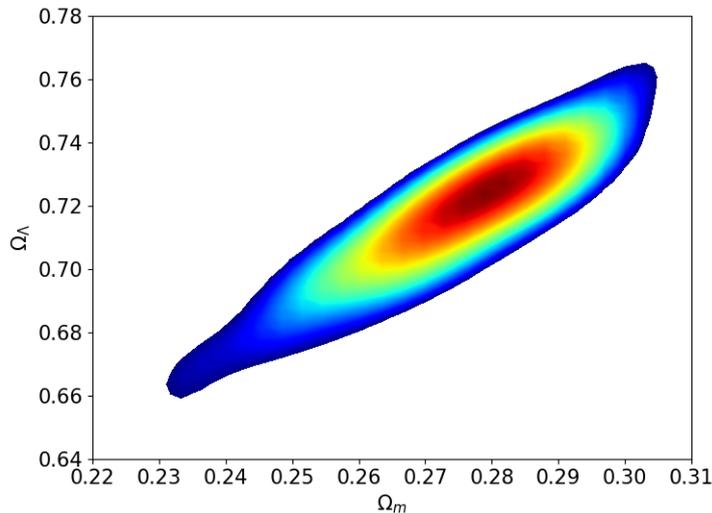
Figure 1.2: Probability density function of an MCMC output for the $\Omega_m \times \Omega_\Lambda$ space. The minimum contour value is equal to a 5% of the maximum value of the PDF.

1.1.3 | CMB Fluctuations and the Universe Flatness

The main evidence for a flat universe comes from the analysis of the temperature fluctuations of the Cosmic Microwave Background (CMB). The CMB corresponds to photons at the epoch of recombination ($z \sim 1000$). It is radiation that permeates the full sky, and its spectral energy distribution is almost a perfect blackbody with a temperature of $T = 2.725K$. Although it appears to be homogenous after some calibrations, the CMB presents temperature fluctuations of the order of $\Delta T/T \sim 10^{-5}$. Therefore, we require high-sensibility detectors to unveil the

---

[2] http://supernova.lbl.gov/Union/figures/SCPUnion2.1_mu_vs_z.txt, which corresponds to a compilation of SNe Ia distance modulus for host galaxies at $0.015 \lesssim z \lesssim 1.4$.

inhomogeneities. The COBE satellite obtained for the first time sensitive enough observations of the full-sky in 1992, revealing the expected anisotropies in the CMB. Many mechanisms produce the CMB fluctuations, and we can divide them into two main classes: *primary* and *secondary* anisotropies. We link the primary anisotropies with inhomogeneities at the epoch when the CMB photons originated, while the secondary is mostly the scattering of the CMB photons along their path to us. Therefore, the origin of CMB fluctuations depend on the primordial structures, the composition, and also the geometry of the universe. We can characterize the CMB anisotropies at different scales similarly to what we explained in Section 1.1.1. The power spectrum of temperature fluctuations is usually given by $l(1 + l)C_l$, that describes the fluctuation amplitudes at angular scales $\theta = \pi/l$. We show in Figure 1.3 the cosmological parameter dependence of the model CMB power spectrum.
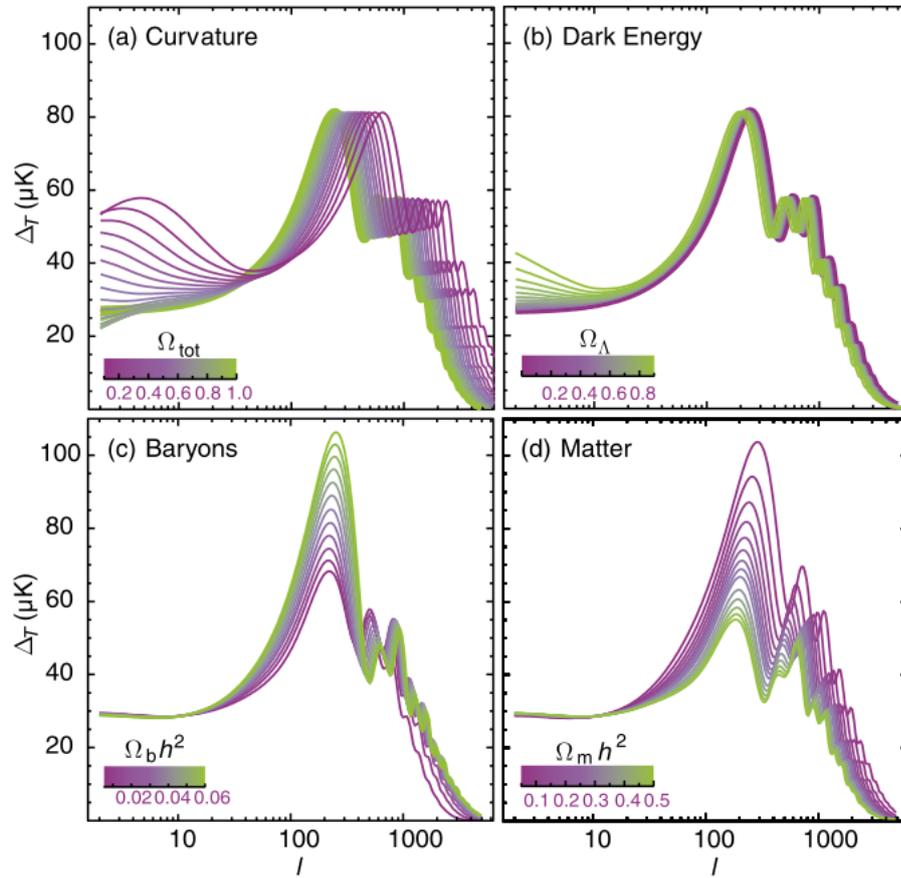
Figure 1.3: The theoretical power spectrum of temperature fluctuations of the CMB and its dependence on the curvature ($\Omega_\Lambda + \Omega_m$; a), dark energy ($\Omega_\Lambda$; b), baryons ($\Omega_b$; c), and matter ($\Omega_m$; d). Figure from Schneider (2006).

As we can see in Figure 1.3, the CMB power spectrum presents a strong dependence on the curvature, baryons, and matter densities. Therefore, the best fit of the observed CMB power spectrum will bring the most likely values for these cosmological parameters with more precision. In particular, the curvature (*a* panel, Figure 1.3) is the most constrained parameter, since the position of the peaks suffer a shift to smaller scales (higher values of l) for a universe with smaller values of $\Omega_\Lambda + \Omega_m$, that implies in a more positive curvature, i.e., higher values of Ω_k, since $\Omega_k = 1 - (\Omega_\Lambda + \Omega_m)$.

Both theoretical and observed power spectrums are consistent with each other; for example, the Planck satellite obtained a χ^2 equals to 1.03 for the best model fit (Planck Collaboration et al., 2014) for a flat universe with $\Omega_m = 0.315$ and $\Omega_\Lambda = 0.685$.

The comparison between the theoretical and observed CMB power spectrum is the main evidence of the flat universe model. This result, combined with those previously discussed in this section, is an important piece for our current knowledge about the nature, composition, and geometry of our universe.

1.2 | Structure Evolution

As we mentioned in the last section, the amplitude of the inhomogeneities at the recombination epoch is of the order of $\Delta T/T \sim 10^{-5}$. On the other hand, present-day galaxy clusters can enclose 200 times the field density within a radius of ~ 1.5 Mpc. It implies that the anisotropies evolve with time. In practice, we quantify the amplitude of inhomogeneities by the *density contrast*, δ:

$$\delta(\vec{r}, t) \equiv \frac{\rho(\vec{r}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{1.5}$$

where $\bar{\rho}(t)$ denotes the mean cosmic density at the epoch t. The gravitational force is the main responsible for the growth of overdensities. For example, in a region with $\delta > 0$, the gravity acts as a break of the expansion of the Universe. Due to it the scale of the overdense structure increases slower than the Universe expansion. Then, δ increases through time. Oppositely, in an underdense region, the gravitational potential is weaker than the mean, implying that its scale grows faster than the universe expansion. Hence, δ decreases. In both cases, the amplitude of the inhomogeneities, $|\delta|$, increases. There are two ways to quantify the growth of the structures. There are two main regimes in the evolution of the structures: the linear and the non-linear. Here, we will present the classic approach for both regimes.

1.2.1 | The Linear Regime

We can assume that the matter in the universe is a pressure-free fluid with density, $\rho(\vec{r}, t)$ and a velocity field, $\vec{v}(\vec{r}, t)$. This approximation is only valid when $|\delta| \ll 1$, which is enough to describe the evolution of structures before the recombination epoch. The behavior of this fluid can be described by the continuity (changes in density due to the velocity field), Euler (influence of forces in the velocity field), and Poisson equations. We present them in Equation 1.6 in comoving coordinates. These coordinates are useful in this type of analysis because if under only

the expansion of the universe, the position of the particles do not change. The position of a particle in comoving coordinates is given by $\vec{x} = \vec{r}/a(t)$, where a is the *scale factor*. Additionally, the matter particles present motions in comoving coordinates that we call peculiar velocities, denoted as $\vec{u}(\vec{x}, t)$. We also express these equations as a function of the density contrast, δ, and the mean cosmic density, $\bar{\rho}$, since we express the density as $\rho = \bar{\rho}(1 + \delta)$. The hydrodynamical equations in this case become:

$$
\begin{aligned}
\text{Continuity}: \quad & \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 & \longrightarrow \quad & \frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta)\vec{u}] = 0 \\
\text{Euler}: \quad & \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} = -\nabla \Phi & \longrightarrow \quad & \frac{\partial \vec{u}}{\partial t} + \frac{\vec{u} \cdot \nabla}{a}\vec{u} + \frac{\dot{a}}{a}\vec{u} = -\frac{1}{a}\nabla \phi \\
\text{Poisson}: \quad & \nabla^2 \Phi = 4\pi G \rho & \longrightarrow \quad & \nabla^2 \phi(\vec{x}, t) = 4\pi G a^2 \bar{\rho} \delta(\vec{x}, t)
\end{aligned}
\tag{1.6}
$$

In the linear regime, we discard all terms of second-order or more in Equation 1.6. Then, the terms $\delta\vec{u}$ and $\vec{u}(\vec{u} \cdot \nabla/a)$ are removed from the continuity and Euler equations, respectively. This approximation leads us to a single differential equation that describes the evolution of the density contrast over time. It can be expressed as a function of the Hubble ($H = \dot{a}/a$) and the matter density parameters, Ω_m ($8\pi G\bar{\rho}/3H^2$):

$$
\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} - \frac{3}{2}\Omega_m H^2 \delta = 0
\tag{1.7}
$$

From Equation 1.7, we can conclude that in the linear regime, the evolution of the inhomogeneities depends only on the cosmological model. Additionally, since Equation 1.7 does not present derivatives in spatial coordinates, the density fluctuations do not suffer alteration of their scales in the comoving frame. Therefore, the structure increases its size following the universe expansion. In an Einstein-de Sitter universe ($\Omega_m = 1$, $\Omega_\Lambda = 0$), the Hubble constant is $H = 2/(3t)$. Then, we obtain that the density contrast evolution in the linear regime is given by the Equation 1.8:

$$
\delta(t) = \delta_0 t^{2/3}
\tag{1.8}
$$

1.2.2 | The Non-Linear Regime

We can obtain expressions of the non-linear theory from Equations 1.6. However, the fluid approximation is not valid if $\delta \gtrsim 1$, because the behavior of this fluid would generate multiple steams of matter, during the formation of gravitationally bound structures (Schneider, 2006). Nevertheless, the top-hat model is a classical example of a theory that provides an analytical solution for the non-linear regime.

Consider a slightly overdense sphere with a mass M within a radius r in a homogeneous and isotropic space. Following the Birkhoff theorem, this spherical perturbation evolves indepen-

dently of the outside. Therefore, we can describe our system using Equation 1.9:

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 - \frac{GM}{r} = E \tag{1.9}$$

Notice that if $E < 0$, the sphere will increase its radius until reaching a maximum, r_m at t_m. We call this moment as *turn-around*. After, the sphere starts to collapse. The solution of Equation 1.9 is obtained in a parametric way. We have that $r/r_m = (1 - \cos\theta)/2$ and $t/t_m = (\theta - \sin\theta)/\pi$ with $0 \leqslant \theta \leqslant 2\pi$, for an Einstein-de Sitter universe. Finally, the expression for the density contrast evolution of the sphere is given by Equation 1.10:

$$\delta = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3} - 1 \tag{1.10}$$
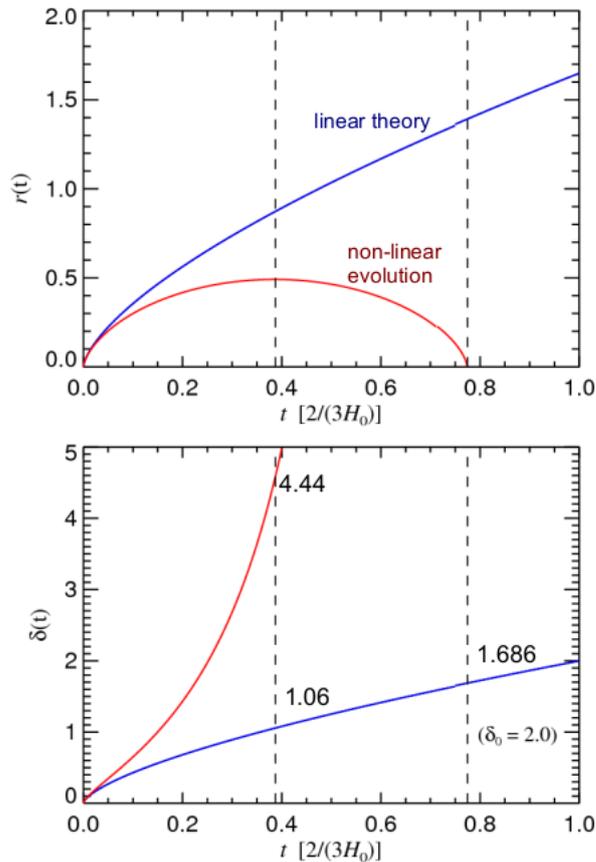
Figure 1.4: Evolution of the scale (top panel) and density contrast (bottom panel) of a slightly overdense sphere in an Einstein-de Sitter universe ($\Omega_m = 1$, $\Omega_\Lambda = 0$). Blue and red lines indicate the predictions for the linear and non-linear regimes. Figure extracted from a Volker Springel lecture https://www.mpi-hd.mpg.de/lin/events/isapp2011/pages/lectures/Springel-1_2.pdf.

We compare the evolution of the scale (top panel) and the density contrast (bottom panel) of an overdense sphere in Figure 1.4 for an Einstein-de Sitter universe for both linear and non-linear regimes.

For early times (small values of t), both approximations are consistent with each other. At the *turn-around* time, the expected density contrast are ~ 1.06 and ~ 4.44 for linear and non-linear

models, respectively.

Notice that at $\theta = 2\pi$, the density contrast of the non-linear model goes to the infinity. After the collapse, the structures will attain an equilibrium state and virialize. From the virial theorem, we can demonstrate that the radius of the virialized structure is half of r_m. Therefore, the density of the system is 8-times the density at t_m. For an Einstein-de Sitter universe, we find the expression in Equation 1.11, since $a_{EdS} \propto t^{2/3}$ and $\rho_{\bar{E}dS} \propto a^{-3}$:

$$\rho_v = 177.7 \rho_{\bar{E}dS}(t_0)(1 + z_v)^3 \tag{1.11}$$

where t_0 is the age of the universe and z_v the redshift when the virialization happened. After the virialization, the density contrast is constant, $\delta_v \sim 176.7$ for the Einstein-de Sitter universe. For other cosmological models, the density contrast for the top-hat model presents a dependence on the redshift of the virialization.

We can write Equation 1.11 as a function of the mass, M, and velocity dispersion, σ, of the system. Hence, we obtain an expression that predicts the epoch when structures, with these characteristics, have formed:

$$1 + z_v \approx 8.8 \left(\frac{\sigma}{100 \text{ km/s}}\right)^2 \left(\frac{M}{10^{12} \ M_\odot}\right)^{-2/3} h^{-2/3} \tag{1.12}$$

Table 1.1 lists the typical formation redshifts of objects with a certain mass and velocity dispersion.

Table 1.1: Formation redshift z_v for three differnt objects with a given mass, M and velocity dispersion, σ. These values are obtained for an Einstein-de Sitter universe.

$1 + z_v$	$M/M_\odot$	σ [km/s]	Object
17	10^{10}	30	Galaxy
8	10^{12}	100	Galaxy
1	10^{15}	500	Galaxy Cluster

Results from Equation 1.12 motivated the idea of the hierarchical model of structure formation (Blumenthal et al., 1984; Davis et al., 1985; Cole et al., 2000) because low mass objects are virialized (formed) first.

The hierarchical model of structure formation suggests that the massive structures (for example, a galaxy cluster of $M = 10^{15} \ M_\odot$) that we observe in the nearby universe have formed from the merger of smaller structures. If we call halos the virialized objects, we expect many smaller halos existed independently at high redshift. We present a scheme of a merger tree in Figure 1.5. In this way, the progenitors of the galaxy clusters, that we call protoclusters, consist of all these small halos that will merge and reach a mass $M > 10^{14} M_\odot$ at $z = 0$ or before.
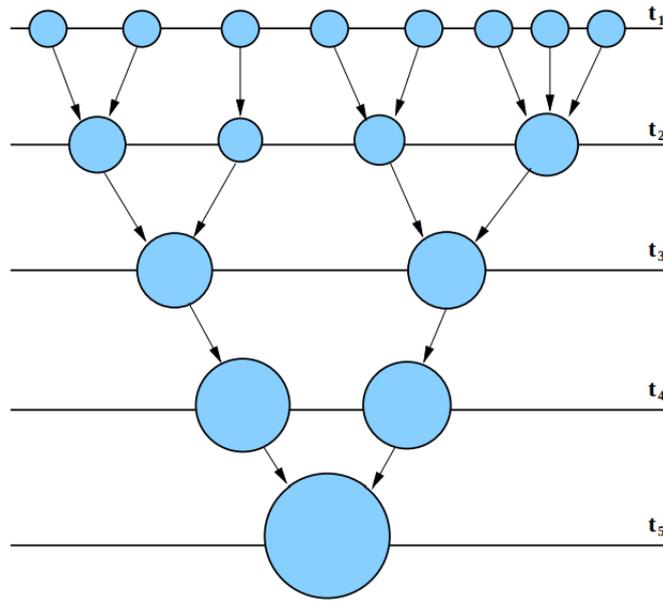
Figure 1.5: A schematic of a merger tree. Horizontal lines represents different epochs with $t_1 < t_5$. The size of the circles are proportional to the dark matter halo mass. Figure from Baugh (2006).

1.3 | Simulations and The Evolution of the Large Scale Structure

A more accurate description of the evolution of the structures as well of the large scale structure of the universe in the frame of the ΛCDM cosmological model can be obtain with numerical simulations.

There are two main types of cosmological numerical simulations: pure dark matter and hydrodynamical simulations. The former, as its name says, predicts the evolution of dark matter only, with the absence of baryons (examples include Millennium Run (Springel, 2005), Bolshoi (Klypin et al., 2011), MultiDark (Prada et al., 2012), among others), while the later simulates both (the most famous ones are: Illustris (Vogelsberger et al., 2014), Horizon-AGN (Dubois et al., 2014), and EAGLE (Schaye et al., 2015)). Contrarily to hydrodynamical simulations that can predict galaxies and their morphology, we cannot compare results from a pure dark matter simulation directly with observational data. However, the dark matter halo merger-trees are the base of the semi-analytical models (SAMs). A SAM corresponds to a set of differential equations that describes the evolution of baryonic components such as cold and hot gas, supermassive black holes, stars, galaxy disks, and bulges, among others. These equations depend on free parameters constrained from observational data.

The two types of numerical simulations demand significant computational work. They have to deal with an enormous quantity of simulation components, which we call particles. A higher number of particles implies in a better resolution. The amount of particles depends on the computational capabilities, and this is one of the limitations of simulations. The simulation

algorithm implements the equations that can describe the evolution of each component in a fixed comoving volume. It saves the state of the simulated universe at different epochs. The characteristics of the particles depend on the number of particles, comoving box size, and cosmological parameters. The comoving box depends on the scientific goal of the simulation. For example, the Millennium Run (MR) simulation (used in this work) (Springel, 2005) had the main goal to track the evolution of the large scale structure of the universe. Therefore, the chosen volume was $(500 \text{ Mpc}/h)^3$. This simulation is a pure dark matter N-body simulation, with $N = (2160)^3$ particles all with the same mass. The mass distribution evolves in a way that depends of the cosmological parameters and the initial conditions.
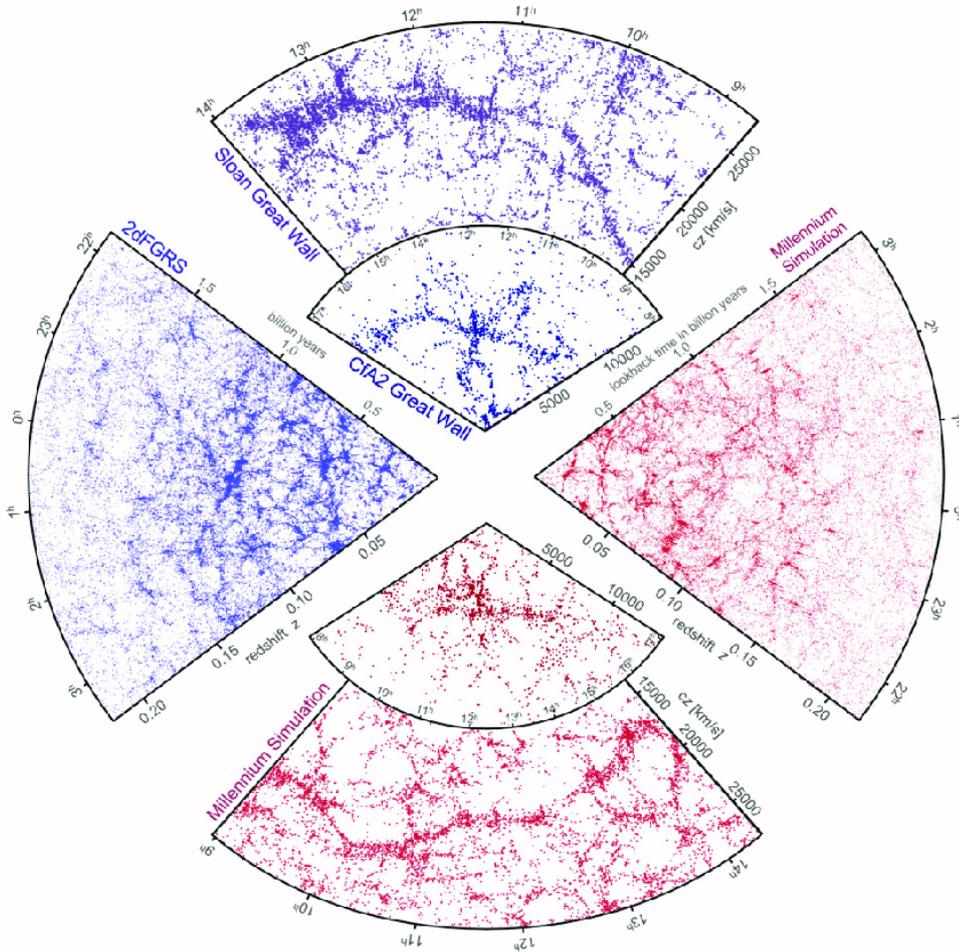
Figure 1.6: Comparison between the galaxy distribution from 2dFGRS and SDSS spectroscopic surveys with a lightcone constructed from the Millennium Run simulation. Figure from Springel et al. (2006).

The MR tracked the evolution of the universe from $z = 127$ to $z = 0$, and it can reproduce satisfactorily the morphology of the large scale structure (LSS), which we call the cosmic web. Figure 1.6 compares the predicted galaxy distribution of the MR simulation plus a SAM with two spectroscopic redshift surveys, the Sloan Digital Sky Survey (York et al., 2000) and the 2-degree Field Galaxy Redshift Survey (Colless et al., 2001). The MR simulation does not only qualitatively reproduced the morphology of the LSS, but also, the observed power spectrum for the magnitude limit of the surveys (Springel et al., 2005).

Nowadays, we cannot perform the comparison shown in Figure 1.6 at high redshifts ($z \gtrsim 1.0$)
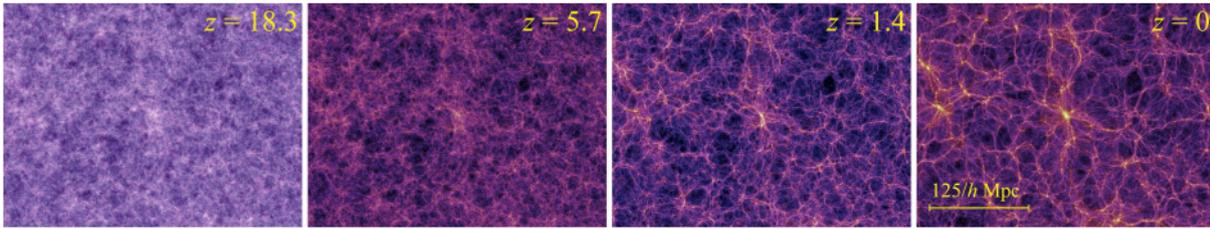
Figure 1.7: Large Scale Structure of the universe at z=18.3, 5.7, 1.4 and 0.0 from the Millennium Run simulation. Figure from Lima Neto et al. (2014).

due to observational constraints. There are several high-redshift spectroscopic surveys, but these are incomplete because it is challenging to obtain spectra of high redshifts galaxies since these are faint. Therefore, a significant fraction of our knowledge about the evolutionary state of the LSS at high redshift comes from simulated data. Figure 1.7 shows the simulated LSS at four different epochs. As we can see, the universe was more homogeneous at early times, and the amplitude of the density contrast grows with time. As we mention in the last section, the scale of the underdense regions increases faster than the universe expansion, due to the weaker gravitational potential compared with the average. The evolution of the overdense regions at the primordial nodes of the cosmic web results in the formation of the largest gravitationally bound structures that we know, the galaxy clusters. The dense central region in Figure 1.7 becomes more compact and overdense with time. However, at higher redshifts, $z \gtrsim 2.0$, we find the galaxy cluster progenitors as a set of independent dark matter halos, the so-called protoclusters. These are the structures that we will focus on this work (see the next section).

Another noticeable characteristic of LSS is the filamentary pattern. The filaments that connect the galaxy clusters in the cosmic web are not discernible at high redshift, and we start to distinguish them at $z \sim 1.4$. This is the expected epoch when a reasonable fraction of galaxy clusters ($M \geqslant 10^{14} \ M_\odot$) have already formed as Figure 1.8 shows.

We can quantify the evolution of the structures in the universe from the differential halo number density as a function of mass and redshift (see Figure 1.8). Press and Schechter (1974) found, for the first time, an expression that links cosmological parameters with the mass distribution of dark matter halos at a specific redshift, assuming a spherical collapse. This function, called Press-Schechter, was extensively used to constrain cosmological parameters. However, this classic analytic mass is not consistent with numerical simulation results. Nevertheless, more recent mass functions (e.g., Jenkins et al., 2001) present a strong consistency with the MR simulation for halos at $z \leqslant 10.07$.
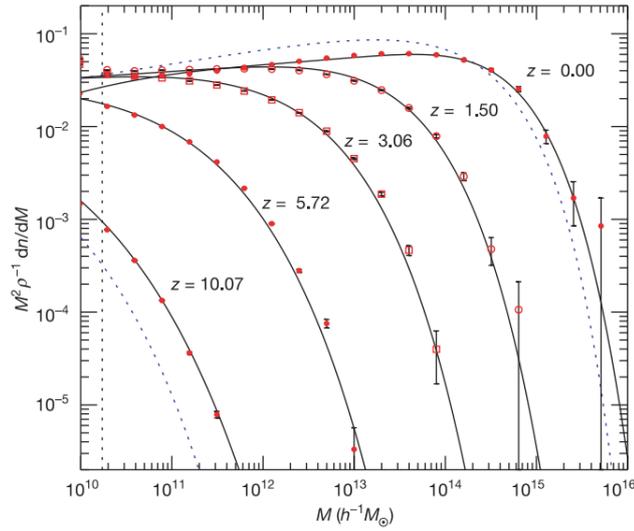
Figure 1.8: Differential halo number density as a function of mass and epoch. Red dots are the mass distribution from the MR simulation with error bars equals to 1σ. Solid black lines are the analytical mass function of Jenkins et al. (2001),while the blue dotted lines at $z = 0.0$ and $z = 10.07$ are the Press-Schechter model. Figure from Springel et al. (2005)

1.4 | Protoclusters

In this work, we focus on the detection of protoclusters in photometric surveys. In this section, we will describe the structures that we are interested in, and their main difference with their descendants, the galaxy clusters.

From the hierarchical model of structure formation, we expect to find protoclusters as an ensemble of independent dark matter halos at high redshifts. Overzier (2016) proposed a formal definition of what a protocluster is: any region at high redshift that will evolve into a galaxy cluster at $z = 0$ or before (see Figure 1.9 for a representation). Before continuing the protocluster description, we will briefly describe the galaxy clusters.
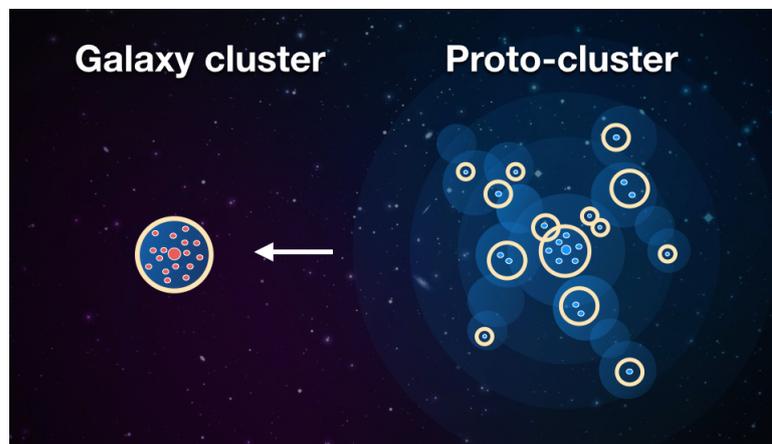
Figure 1.9: A schematic protocluster representation. Figure from Yi-Kuan Chiang profile in University of Texas at Austin website (currently, it is offline).

1.4.1 | Galaxy Clusters

Galaxy clusters are the biggest structures in equilibrium (or in quasi-equilibrium) in the universe, located at the nodes of the cosmic web. They present high masses, above $10^{14} M_\odot$ (Vikhlinin et al., 2014), and large galaxy overdensities (~ 200) within a radius of ~ 1.0 Mpc. The principal mass component is the dark matter. They present a conspicuous Intra-Cluster Medium (ICM) occupied by a hot gas emitting X-rays with temperatures between $T \sim 10^7$ and 10^8 K. Most of the baryonic mass in galaxy clusters is in this hot gas. It also produces fluctuations in the Cosmic Microwave Background (CMB), known as the Sunyaev-Z'eldovich (SZ) effect, where the involved physical process is the inverse Compton scattering, where free electrons interact with the CMB photons, shifting them to higher frequencies. These structures, living in (quasi-) virialized halos, frequently have their brightest galaxies at their centers which, in most cases, coincide with the peak of their X-Ray emission (Lin and Mohr, 2004). Clusters also have a large population of early-type galaxies showing a well-defined "red sequence" in color-magnitude diagrams (Gladders et al., 1998).

These characteristics make clusters perfect laboratories to study effects on galaxy evolution due to their environments. High galaxy overdensities favor the interaction between them. However, in these systems, the members present high velocities that inhibit the mergers often seen in poor groups. Nevertheless, galaxy harassment (Moore et al., 1998), the cumulative effect of high velocity encounters, affects galaxy morphology. Another galaxy-galaxy interaction in clusters is the cannibalism: the most massive galaxy, in general located at the center, affects the trajectory of the other members due to the dynamical friction caused by their movement in this dense environment. This generates that less massive galaxies start to move towards the most one. The ICM also plays an important role: when a galaxy infalls into the cluster, the hot ionized gas removes most of its available gas through a process called ram-pressure streaming (Gunn and Gott, 1972). Additionally, there is evidence that many galaxies quenched their star formation when they were still members of infalling groups. This is known as pre-processing (Zabludoff and Mulchaey, 1998), which explains the significant fraction of passive galaxies beyond the galaxy cluster virial radius (Hou et al., 2014; Bianconi et al., 2018; Pallero et al., 2019).

The effect of all these processes reflects on the galaxy population. Examples include, the morphology-density relation (Dressler, 1980), where the fraction of elliptical and lenticular galaxies increases with the local density, whereas the fraction of spirals decrease, and also the Butcher-Oelmer effect (Butcher and Oemler, 1978) that suggests that at higher redshifts, the galaxy population was more star-forming compared to those in the present-day.

There are three main ways to detect galaxy clusters in three different spectral ranges of the electromagnetic spectrum. We can observe the ICM gas in X-rays, due to their extremely high temperatures. Therefore, searching for extended X-ray sources is a common practice to detect cluster candidates (e.g., Ebeling et al., 2000, 2001; Böhringer et al., 2004). Additionally, we can derive the electron density and temperature profiles. Assuming hydrostatic equilibrium, we can estimate the cluster mass. Galaxy clusters are also associated with extended sources in microwave observations due to the SZ effect (e.g., Staniszewski et al., 2009; Planck Collaboration et al., 2016)

but, in this opportunity, we detect the scatter of the CMB photons, instead of observing the ICM hot gas by itself. From SZ detections, we also can derive the cluster mass. Finally, we can detect galaxy clusters from optical photometry, by analyzing the spatial distribution of galaxies (e.g., Gladders and Yee, 2005; Wen and Han, 2011; Durret et al., 2011). In general, a common technique is by using photometric redshifts to select galaxies within a redshift slab, whose width is chosen according to the photo-z accuracy. Then, a density estimator (or clustering) algorithm is applied to the galaxy sample within $z \pm \Delta z$ and, we associate overdensity peaks with cluster candidates. Another method, based on the expected galaxy population in clusters (Rykoff et al., 2014), i.e., works with galaxy colors, instead of in the (photometric) redshift space.

Although spectroscopic confirmation of cluster candidates is required, the sample of confirmed $z < 1$ clusters is quite large. This is not the case for protoclusters due to observational constraints and physical properties of these high redshift structures.

1.4.2 | Detecting Protoclusters

Although we can detect signals of already formed hot ICM at $1.0 < z < 2.0$ (e.g., Santos et al. (2011) and references therein), we only observed this characteristic in the most massive structures at high redshift. Therefore traditional techniques, based on ICM, are no longer effective to detect protoclusters. Additionally, contrarily to what we observe in the local universe, where clusters host mostly passive galaxies, many star-forming galaxies and AGN activity is observed in protoclusters (e.g., Overzier et al., 2006; Boris et al., 2007; Koyama et al., 2013). However, this decade has seen a significant progress in the study of structures at high redshift. New protoclusters and high redshift cluster candidates are being continuously reported from the analysis of surveys, mostly in the optical and infrared. For example, Chiang et al. (2014) reported 36 photometric redshift selected protoclusters at $z \sim 1.5 - 3.0$ in the COSMOS field. Toshikawa et al. (2016) cataloged 21 protocluster candidates in the Deep survey of CFHTLS using the dropout technique; applying the same method over the wide layer of HSC-SSP area, Toshikawa et al. (2018) presented 210 new candidates at $z \sim 4.0$. Also, Martinache et al. (2018) selected 2151 and 228 protocluster candidates from *Planck* and *Herschel* fields, respectively. From these, 89 were observed with *Spitzer*/IRAC, with more than 92% of them presenting significance overdensities. More recently, Gonzalez et al. (2019) have found a total of 1787 high redshift cluster candidates in the sky covered by the Pan-STARRS and SuperCOSMOS surveys.

Another observational approach targets certain types of objects assuming that they probe dense regions. Examples include radio-loud active nuclei (e.g. Overzier et al., 2006; Venemans et al., 2007; Hatch et al., 2011b,a; Hayashi et al., 2012; Wylezalek et al., 2013; Cooke et al., 2014; Hatch et al., 2014; Cooke et al., 2016), supported by simulations by Orsi et al. (2016) and Lovell et al. (2018). At the same time, despite optically selected quasars being often used to trace overdense regions (e.g. Boris et al., 2007; Adams et al., 2015; Onoue et al., 2018; Stott et al., 2020), the population of quasars generally does not appear to probe proto-clusters (Champagne

et al., 2018; Uchiyama et al., 2018; Vicentin et al., submitted), and Yoon et al. (2019) suggest that massive galaxies are better tracers of overdense regions than quasars.

A common approach to find high redshift structures is by adopting colour criteria using narrow-band or broad-band filters. With the former, it is possible to select Hα or Lyα emitters (hereafter, HAEs and LAEs, respectively), and search for an excess of these objects at similar redshifts (e.g., Shimakawa et al. (2018a,b) for HAEs) and (e.g., Venemans et al. (2005, 2007); Chiang et al. (2015); Higuchi et al. (2019) for LAEs). Similarly, Toshikawa et al. (2016, 2018) found a large number of potential protoclusters from the clustering of Lyman Break Galaxies selected through the dropout technique. Figure 1.10 shows examples of protocluster candidates from Toshikawa et al. (2018).
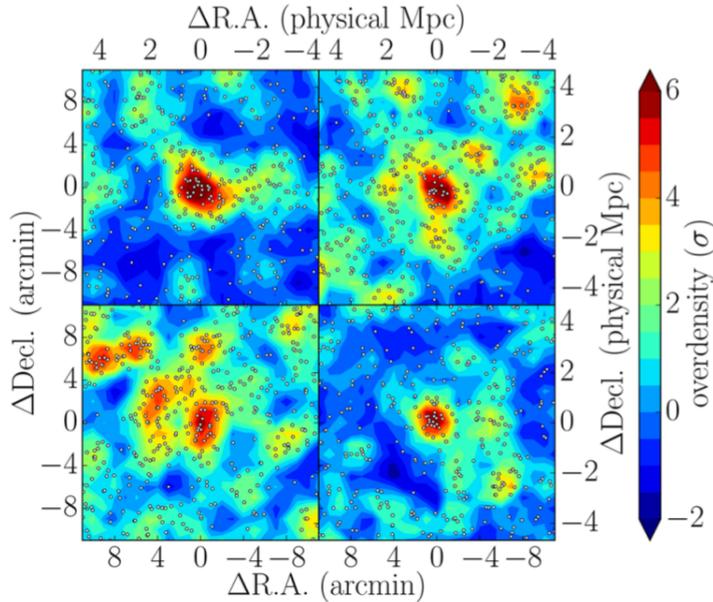
Figure 1.10: Example of protocluster candidates from Toshikawa et al. (2018). White dots mark the position of the *g-dropouts*.

Spectroscopic follow-up of the member candidates is required to confirm if the protoclusters are indeed real but, since high redshift sources present faint magnitudes, it is challenging to obtain their spectra. Additionally, candidate protocluster members are interesting targets for spectroscopic surveys designed to explore galaxy evolution. Examples include surveys such as z-COSMOS (Lilly et al., 2007, 2009), the VIMOS Ultra Deep Survey (VUDS) (Le Fèvre et al., 2015), Gemini Observations of Galaxies in Rich Early Environments (GOGREEN) (Balogh et al., 2017), among others. In particular, the future Prime Focus Spectrograph (Takada et al., 2014) will be able to confirm a large number of protocluster candidates between $1.0 \lesssim z \lesssim 8.0$.

1.4.3 | Studying Protoclusters

A large variety of sources can reside in protoclusters, such as Lyman Break Galaxies, Lyman α emitters, Lyα blobs, Hα emitter. In some cases, we find quasars, radio-loud quasars, gas-rich galaxies, and extremely star-forming galaxies. Then, the study of the galaxy populations in protoclusters can give us important clues to a better understanding of galaxy evolution.

Some systems are expected to evolve into the biggest structures currently known in the local universe, making these protoclusters particularly interesting targets to study the first stages of galaxy assembly in dense regions. For example, Steidel et al. (1998) reported the SSA22 protocluster, a structure at $z \sim 3$ of ~ 10 Mpc scale, analyzing the clustering of Lyman Break Galaxies with measured spectroscopic redshift. Pentericci et al. (2000) confirmed a significant quantity of galaxy members in the Spider-web protocluster, which presents the early stage of a typical Brightest Cluster Galaxy (BCG). Miller et al. (2018), with ALMA observations, detected 14 extremely starburst galaxies at $z \sim 4.3$ within a projected radius of 130 kpc (SPT protocluster); this is another example of a proto-BCG. Cucciati et al. (2018), with spectroscopic redshifts from the VIMOS Ultra Deep Survey, traced a particular structure, called Hyperion, composed of at least seven overdensity concentrations. Recently, Long et al. (2020) found a protocluster core at $z \sim 4$, dubbed Distant Red Core (DRC), whose estimated mass is comparable to the Virgo cluster (see Figure 1.11). Simulations do not predict the existence of some of these structures. In this sense, the characteristics of the confirmed protoclusters can unveil the limitations of our current models of structure formation and evolution.
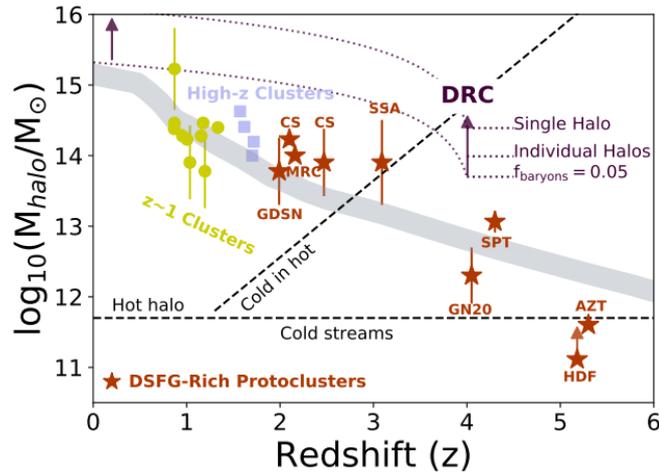
Figure 1.11: Comparison of reported protocluster halo mass (symbols) with simulation predictions (gray shaded area). Figure 7 from Long et al. (2020).

Nevertheless, our knowledge of protoclusters can also be guided by simulations, which can be used to predict the main properties of these structures and their members. For instance, Chiang et al. (2013) have shown that a significant fraction of protocluster mass at $z = 2.0$ is within a radius ~ 4 times larger than the typical comoving radius at $z = 0$ and ~ 6 times larger at $z = 4.0$ (see Figure 1.12). Muldrew et al. (2015) found that just $\sim 20\%$ of the total protocluster mass is

enclosed by the main dark matter halo at the same redshift. Investigating galaxy properties in semi-analytic models, Muldrew et al. (2018) have shown that galaxies in the principal halo have lower star-formation rates than galaxies in smaller halos and that both have lower star formation rates than field galaxies due to the accelerated evolution in these overdense regions. Recently, Trebitsch et al. (2020) using the OBELISK hydrodynamical cosmological simulation, found that stellar populations provide enough energetic photons to complete the HI reionization at $z \sim 6.0$ without other additional ionizing sources (e.g., AGNs and collisions). These results suggest that reionization happens from inside the galaxy overdense region to the entire intergalactic medium.
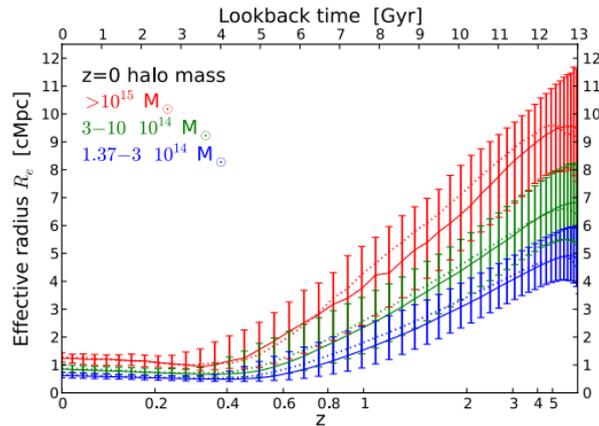
Figure 1.12: Effective radius of protoclusters as a function of redshift. Error bars indicate the 1-σ dispersion. Figure 2 (right panel) in Chiang et al. (2013).

Observational results are commonly interpreted with simulation (e.g. Overzier et al., 2009; Chiang et al., 2013; Toshikawa et al., 2016; Lemaux et al., 2018; Jiang et al., 2018; Shi et al., 2019; Long et al., 2020). In particular, lightcones are often constructed to analyze observational data from simulations that emulates observations (Blaizot et al., 2005; Kitzbichler and White, 2007; Merson et al., 2013; Overzier et al., 2013; Stothert et al., 2018; Araya-Araya et al., submitted) (see Figure 1.13 for a lightcone representation). The usage of lightcones helps to address the purity and completeness of detection or selection of structures (Kim et al., 2016; Ascaso et al., 2016; Costa-Duarte et al., 2018; Euclid Collaboration et al., 2019; Krefting et al., 2020; Werner et al., in prep.), as long as observational constraints are properly taken in to account.

Future projects, such as the Vera Rubin Observatory's Legacy Survey of Space and Time (LSST), Euclid, and the James Webb Space Telescope, as well as the new generation of very large telescopes (Giant Magellan Telescope, Extremely Large Telescope, Thirty Meter Telescope) will open new windows to discoveries. Simulated data has been used to evaluate forecasts for high-redshift surveys (Bisigello et al., 2016; Graham et al., 2018; Laigle et al., 2019; Graham et al., 2020). However, despite the usefulness of lightcones to estimate the significance level of selection criteria and expected detections, some technical aspects of their construction, for example the low completeness of structures at particular redshifts, may difficult the interpretation of observations, due to the poor sampling of the structures in certain redshift ranges with these lightcones.

In this Dissertation, we introduce the protocluster-lightcones, dubbed PCcones, as a tool to help this type of analysis, focusing on photometric redshift surveys using broad-band photometry, and emphasizing structure detection at $1.0 \leqslant z \leqslant 3.0$. It consists of a π deg^2 lightcone
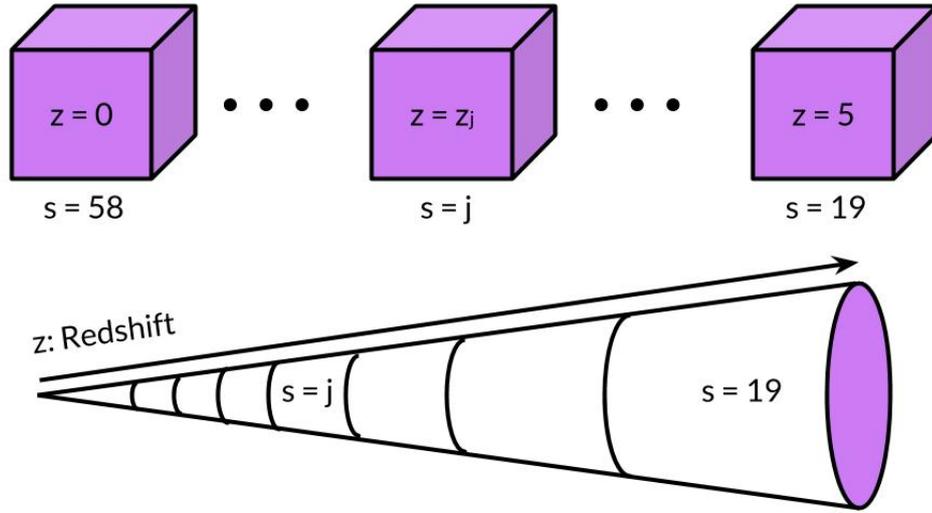
Figure 1.13: Lightcone basis concept scheme: construction of a continuous space in redshift and angular coordinates to recreate an observation from simulated data.

with a pre-selected structure, like a $z = 0$ galaxy cluster progenitor, placed at a desired (high) redshift. This is particularly helpful for investigating protocluster candidate detections from imaging. Our motivation to introduce these PCcones is that we can add observational constraints in the analysis (like limits in magnitudes) and examine their impact in the detection of galaxy overdensities, evaluating the likelihood of detected structures being real, determining detection rates, and estimating the expected quality of photometric redshift selection, allowing to estimate the probability that an observed overdensity for a given magnitude limit and photo-z selection is indeed a protocluster. This approach has several advantages, such as helping to design protocluster surveys, interpret their results, and to justify spectroscopic follow-ups.

In this work, we assume a ΛCDM universe, with cosmological parameters equal to those obtained by the *Planck1* mission results (Planck Collaboration et al., 2014): $h = 0.673$, $\Omega_m = 0.315$ and $\Omega_\Lambda = 0.685$.

2 Protocluster-Lightcones

In this chapter we describe the procedure adopted to construct protocluster lightcones. The results of this dissertation are obtained by placing at different redshifts ($z = 1.0$, 1.5, 2.0, 2.5 and 3.0) 20 cluster progenitors obtained from the Millennium Simulation (Springel, 2005) with $M_{z=0} > 1.37 \times 10^{14} M_\odot$ (see Section 2.8 for more details about the placed protoclusters); this value corresponds to the low mass Fornax-type protoclusters (Chiang et al., 2013). Here $M_{z=0}$ denotes the descendant cluster mass at $z = 0$. We define a galaxy cluster following Chiang et al. (2013), that is, a main Friend of Friends dark matter halo at $z = 0$ with `M_tophat` $> 1 \times 10^{14} M_\odot$ (notice that the $z = 0$ snapshot in the *Planck1* scaled catalog is not the 64*th* snapshot; it is, instead, the 58*th*).

2.1 | Synthetic Galaxies

The first step in the construction of lightcones is to obtain simulated galaxies. In this work, we used the public release of the `L-GALAXIES` semi-analytic model (Henriques et al., 2015). In short, a semi-analytic model (SAM) corresponds to a set of differential equations describing the evolution of the baryonic components (the dark matter is simulated directly using a N-body code, in this case the Millennium Simulation); the model has free parameters that can be constrained by observations. The differential equations act over primordial gas associated to dark matter particles set at the beginning of a dark matter N-body simulation. The primordial gas evolves and is transformed into other baryonic components, such as stars, black holes, and intracluster hot gas, depending on the model. The SAM output gives physical properties of the synthetic galaxies, such as stellar mass, gas mass, star formation rate, among others (see Henriques et al., 2015, for more details). An important output of `L-GALAXIES` are the star formation histories arrays, which we will use to estimate spectro-photometric properties of the model galaxies.

We apply the `L-GALAXIES` SAM to the Millennium Run simulation (Springel et al., 2005) scaled to the cosmological parameters obtained from the *Planck1* data (Planck Collaboration et al., 2014); the scaling from the original cosmological parameters to the new ones has been performed using the Angulo and White (2010) algorithm. We use the Millennium simulation because of their dark matter particle mass ($m_p \sim 9.6 \times 10^8 \ M_\odot/h$) and box size ($L = 480.279$ Mpc/h), which allow us to model satisfactorily galaxies with $M_\star > 10^8 \ M_\odot/h$ and contains a reasonable high number of massive (proto)clusters in this volume. Also, we choose this simulation for simplicity, since the already produced merger trees by using the `SUBFIND` algorithm (Springel

et al., 2001) are the basis of L-GALAXIES.

2.2 | Lightcone Space Definition
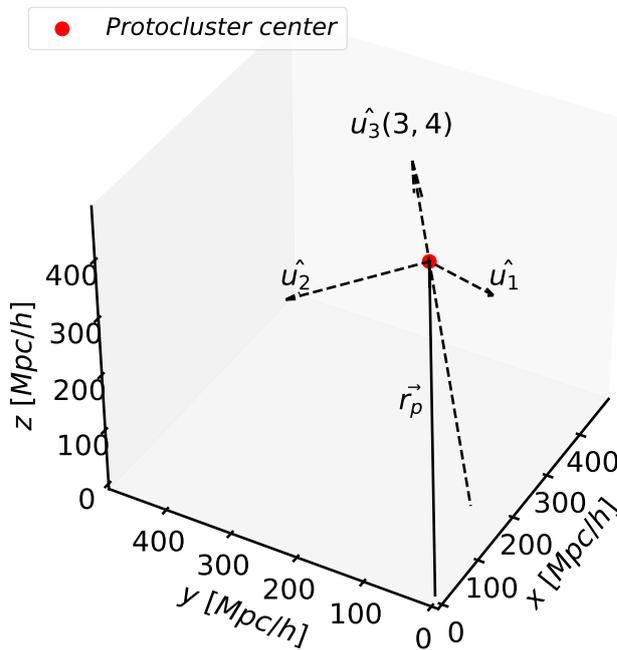
Figure 2.1: Definition of lightcone coordinates, where $\vec{r_p}$ indicates the position of the protocluster that we want to place at z_p in the lightcone. $\hat{u_3}$, $\hat{u_1}$, and $\hat{u_2}$ are orthogonal vectors that define the lightcone space. $\hat{u_3}(3,4)$ represents the line-of-sight with $n = 3$ and $m = 4$.

We want to place the progenitor of a galaxy cluster in the centre of a lightcone and at a redshift $z = z_p$. We define the lightcone space as given by the orthonormal coordinates $(\hat{u_1}, \hat{u_2}, \hat{u_3})$, similar to the procedure explained in Kitzbichler and White (2007). The line-of-sight direction $\hat{u_3}$ is set as $\vec{u_3} = (n, m, nm)$, and we adopt the values, $n = 3$ and $m = 4$. Two additional orthonormal vectors, $\hat{u_1}$ and $\hat{u_2}$, are used to obtain the simulated angular coordinates: right ascension, α, and declination, δ (see Section 2.4).

The progenitor of a galaxy cluster at $z = z_p$ resides in some snapshot of the Millennium Simulation. Since $z = z_p$ is not, in general, the redshift of a Millennium snapshot, we search for the protocluster in the snapshot s_j, with $z_j \leqslant z_p < z_{j+1}$, where z_j and z_{j+1} represent the redshift of the snapshots s_j and s_{j+1}, respectively.

To place the center of mass of the protocluster, $\vec{r_p}$, at the comoving distance $d_C(z_p)$, we re-define the zero point, by putting it at the $z = 0$ position. Then, if $\vec{r_p} = \hat{u_3}\, d_C(z_p) + \vec{r_o}$, we

can establish the position of the zero point, $\vec{r_o}$, by:

$$\vec{r_o} = \vec{r_p} - \hat{u_3}d_C(z_p) \tag{2.1}$$

Figure 2.1 shows an illustration of the lightcone space, with the line-of-sight, the orthonormal vectors, and the protocluster center.

2.3 | Lightcone Volume

Each Millennium snapshot is a box of comoving side $L = 480.279$ Mpc/h, containing the evolutionary stage of the same universe volume at a certain epoch. It is essential to use the information of all available snapshots towards the redshift limit of the lightcone, z_f. We extract galaxies from the snapshot s_j which are located between the comoving distance $d_C(z_j)$ and $d_C(z_{j+1})$, where z_j and z_{j+1}, represent the redshift of the snapshot s_j and s_{j+1}, respectively. Therefore, we need to estimate the position that represents the comoving distance of every snapshot, $\vec{r_s}$, considering the zero-point, $\vec{r_o}$. Using Equation 2.2, we derive the position where the snapshot s_j starts:

$$\vec{r_s}(z_j) = d_C(z_j)\hat{u_3} + \vec{r_o} \tag{2.2}$$

where $d_C(z_j)$ is the comoving distance to the redshift that corresponds to the snapshot s_j, in units of Mpc/h. We construct lightcones of π deg^2, and, consequently, we remove all galaxies outside a radius of 1.0 deg from the line-of-sight $\hat{u_3}$.

Due to the finite size of each snapshot, it is necessary to replicate them. The number of replications depends directly on how deep in redshift the lightcone goes. Therefore, we have created an extended volume which corresponds to a box comprised of $(e - s)^3$ successive Millennium snapshot volumes (e and s explained below). In this new space, galaxy positions can take values between $sL \leqslant x, y, z \leqslant eL$. Volume replication can generate the repetition of certain positions, implying that the same galaxy could appear twice or more in the lightcone, but at different redshifts. Nevertheless, using the values $n = 3$ and $m = 4$ for the line-of-sight direction[1], $\hat{u_3}$, we can avoid the galaxy replication effect out to $z \sim 5.0$ (Kitzbichler and White, 2007).

Since the positions of all SAM output galaxies are stored in x, y and z coordinates $\leqslant L$, we need to transform their coordinates to the extended volume; to do this, we use equation 2.3:

$$\vec{r_c} = \vec{r} - L\vec{i} \tag{2.3}$$

where $\vec{r}$ is the galaxy position in the SAM output and $\vec{i}$ is the 3-D replication index; the $\vec{i}$ vector depends on the replicated volume from where the galaxies were extracted and then placed into the lightcone. The replication index takes integer values in the range $(s, s, s) \leqslant (i_x, i_y, i_z) \leqslant (e, e, e)$, where $\vec{i} = (s, s, s)$ and $\vec{i} = (e, e, e)$ contain the position of the first and last snapshots, respectively.

---

[1] Remember that the line-of-sight direction $\hat{u_3}$ is given by $\vec{u_3} = (n, m, nm)$.

We can then estimate s and e as:

$$s = \texttt{round}\left(\frac{\vec{r_s}(z=0)\cdot \hat{u}_3}{L}\right) - 1$$

$$e = \texttt{round}\left(\frac{\vec{r_s}(z=z_f)\cdot \hat{u}_3}{L}\right)$$

(2.4)

A representation of the lightcone, the extended volumes, and also of the volumes extracted from each snapshot is presented in Figure 2.2.
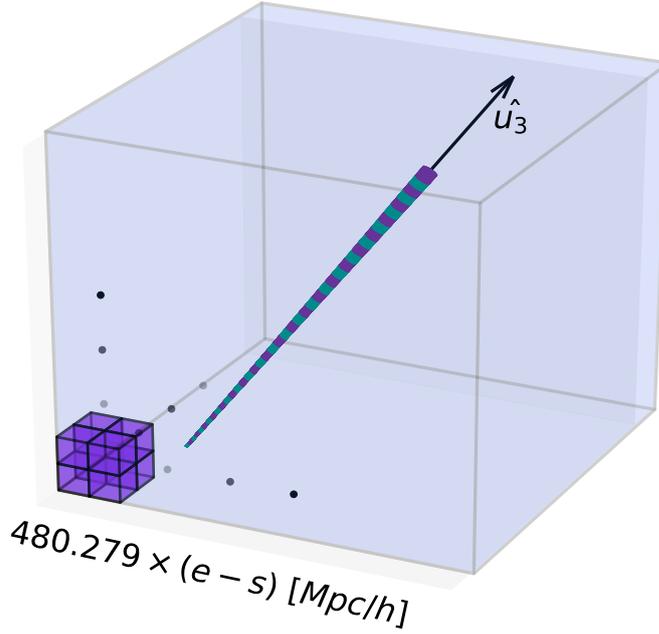
Figure 2.2: The entire lightcone volume, it contains $(e-s)^3$ entire Millennium volumes and this represents a huge volume with side: $(e-s)\times 480.279$ Mpc/h. The Figure also shows the lightcone within the created volume. If the snapshot s_j is an odd or even number, then the volume where we select galaxies is represented in purple or cyan, respectively.

2.4 │ Redshifts and Angular Coordinates

The comoving distance of each galaxy is:

$$d_{C,\text{gal}} = (\vec{r_c} - \vec{r_o})\cdot \hat{u}_3$$

(2.5)

where $\vec{r_c}$ is the galaxy position in the extended volume coordinate system.

For simplicity, the galaxies put in the mock are all those satisfying the condition $d_C(z_j) \leqslant d_{C,\text{gal}} < d_C(z_{j+1})$, where z_j represents the redshift of the snapshot s_j, i.e., those galaxies that are at a comoving distance between the comoving distances of two successive snapshots.

Since the comoving distance is measured along the direction $\hat{u}_3$, to estimate the redshift related to this distance (geometric redshift) we assume that all galaxies with a comoving distance between $d_C(z_k) \leqslant d_{C,\text{gal}} \leqslant d_C(z_k) + 30$ kpc are at $z = z_k$, where z_k takes discrete values between 0 and z_f spaced by the redshift interval equivalent to 30 kpc at z_k. This approach, that we call *Loaf Bread* model, leads to a curvature in the projected sky positions of the galaxies, illustrated in Figure 2.3. Bellow we estimate angular coordinates taking in to account this spherical projection, although this effect is not discernible for small angles ($\lesssim 3$ degrees).
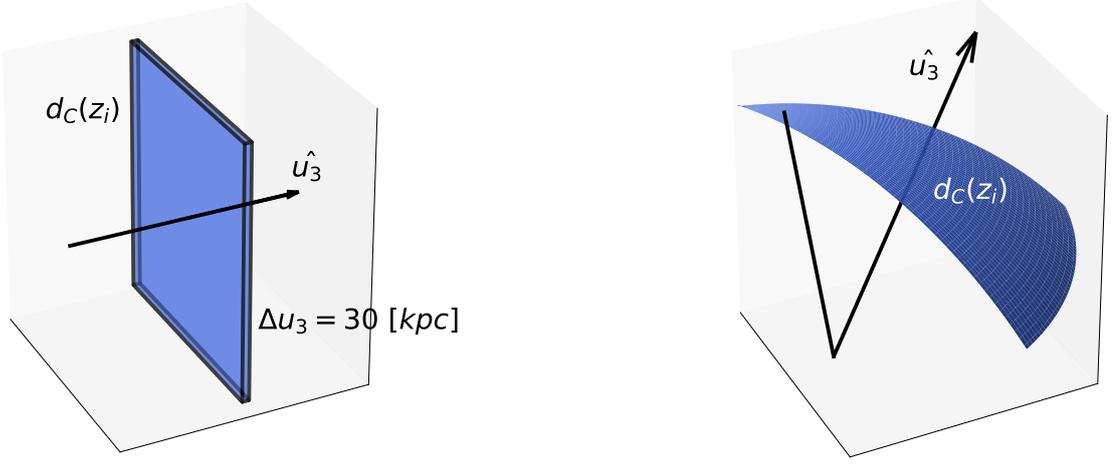
Figure 2.3: Illustration of the geometry of the Loaf Bread Model.

The projection of the position of the galaxies in the $\{\hat{u}_1, \hat{u}_2, \hat{u}_3\}$ space give us the transverse and the radial comoving distances. To estimate the right ascension, α, and declination, δ, we use the projection of the position of each galaxy in the $\hat{u}_1$ and $\hat{u}_2$ directions. Following Kitzbichler and White (2007), α and δ are obtained as

$$
\begin{aligned}
\alpha &= \arctan\left(\frac{\vec{r_c} \cdot \hat{u}_1}{d_{C,\text{gal}}}\right) \\
\delta &= \arctan\left(\frac{\vec{r_c} \cdot \hat{u}_2}{d_{C,\text{gal}}}\right)
\end{aligned}
\tag{2.6}
$$

The next step is to include the peculiar motions, by adding the radial velocities of the galaxies to their geometric redshifts. To apply this correction, we first compute the position of all galaxies in the lightcone space, $\vec{r_l}$:

$$
\vec{r_l} = [\vec{r_c} \cdot \hat{u}_1, \ \vec{r_c} \cdot \hat{u}_2, \ d_{C,\text{gal}}]
\tag{2.7}
$$

and, after, we estimate the radial velocity of a galaxy as $v_r = \hat{r_l} \cdot \vec{v}$, where $\vec{v}$ is the velocity vector in the Millennium simulation at coordinates (x, y, z). The "observed" redshift z_{obs} of each galaxy is then given by

$$
z_{\text{obs}} = (1 + z_{\text{geo}})\left(1 + \frac{v_r}{c}\right) - 1
\tag{2.8}
$$

where z_{geo} is the redshift associated to the comoving galaxy distance, v_r is its radial velocity, and c is the speed of light.

This procedure allows us to obtain the spatial distribution of galaxies in a π deg^2 lightcone,

from $z = 0$ to $z = z_f = 7.0$. Figure 2.4 shows the declination-redshfit distribution between $z = 0$ and $z = 5$ of galaxies brighter than $i = 26.2$ AB mag; this magnitude corresponds to the 5σ magnitude limit of the Wide layer of the Hyper Suprime-Cam Subaru Strategic Program (Aihara et al., 2018). We restrict the mock galaxies right ascension to the interval $\alpha = \pm 0.1$ degrees in Figure 2.4, for better visualization.
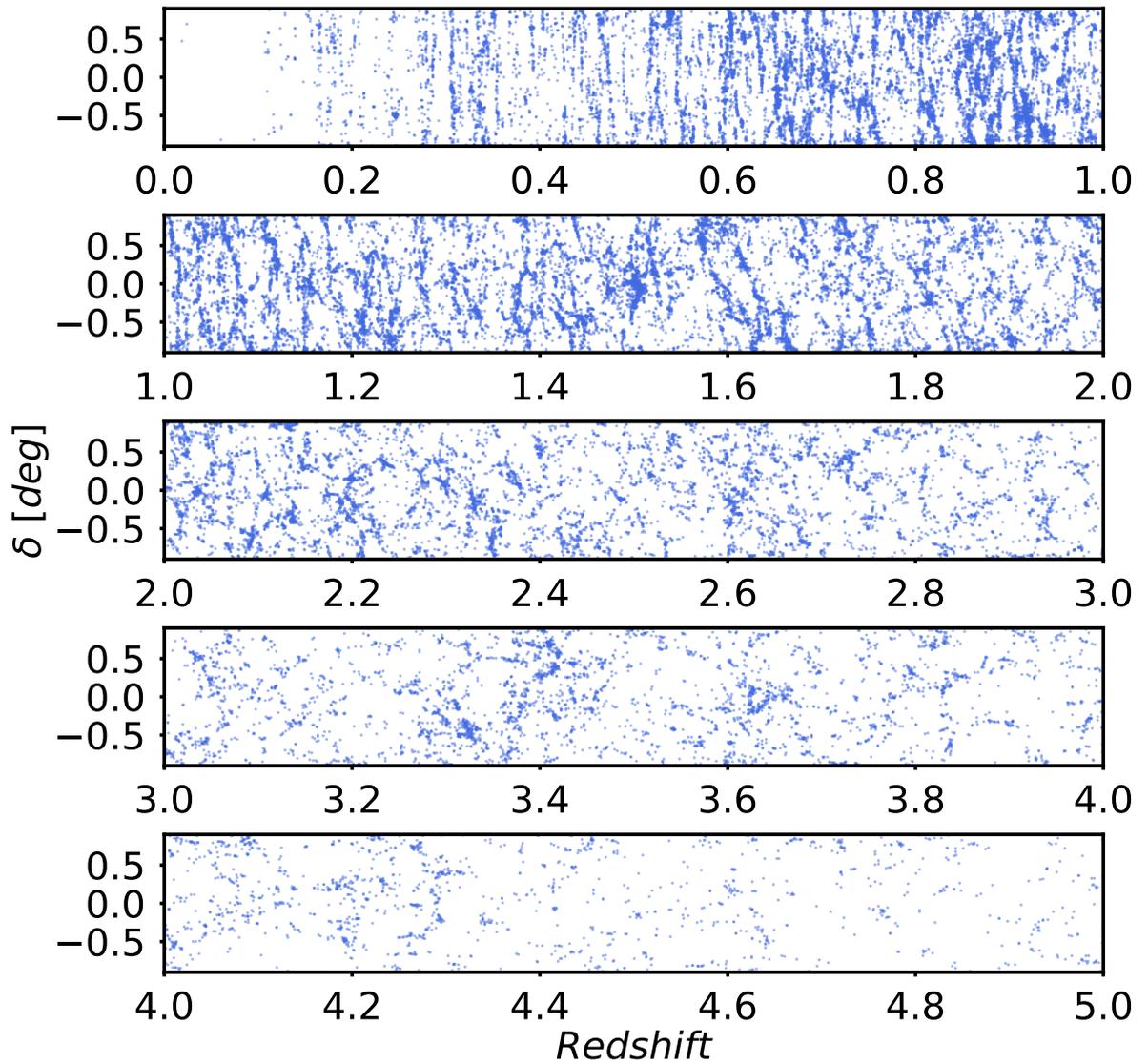
Figure 2.4: Lightcone declination against lightcone redshift. We show simulated galaxies brighter than $i < 26.2$, the nominal magnitude limit of the HSC-SSP Wide survey. The figure presents only galaxies with simulated α between -0.1 and +0.1 degrees.

2.5 | Galaxy SEDs

As mentioned before, we configure the L-GALAXIES SAM to output the star formation history (SFH) of each galaxy. The SFH comprises two arrays, storing the stellar mass and metallicity that was produced between two cosmic times for three different baryonic components: disk, bulge, and intra-cluster medium. We assume that disk and bulge SFH bins represent the stellar populations of each galaxy and, with age, metallicity and stellar mass, we can attribute a spectral energy distribution (SED) for each stellar population in each SFH bin. We present a representation of the SFH arrays in Figure 2.5, where each box is an SFH bin. Notice that the age resolution changes for different snapshots.
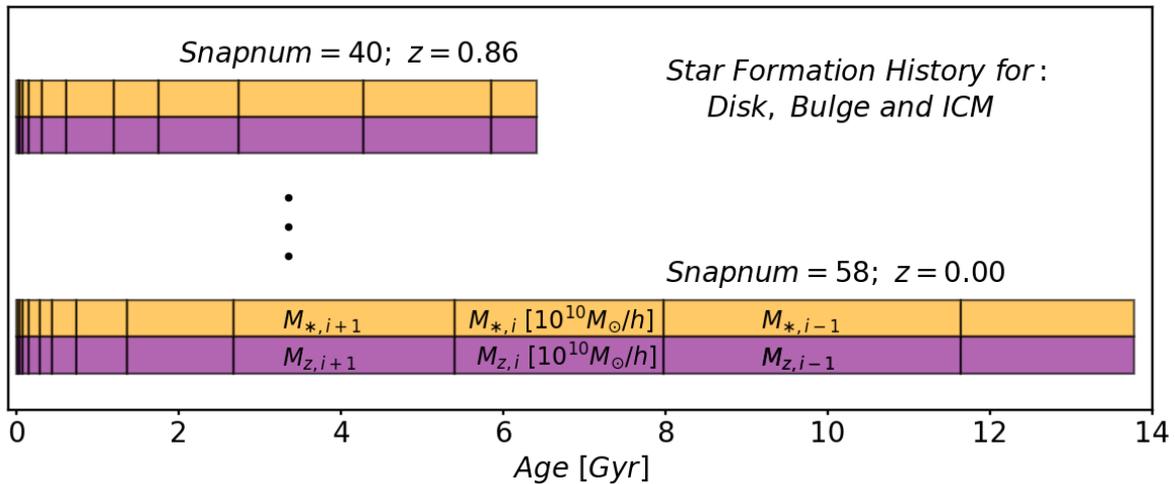
Figure 2.5: Representation of the star formation history arrays for the snapshots 58 and 40. The width of each box indicates the age resolution of each SFH bin, where $M_{*,i}$ and $M_{z,i}$ are the masses in stars and metals produced between the delimited cosmic times of the i-th bin.

The *post-processing* spectro-photometric estimation that we adopted follows Shamshiri et al. (2015): we estimate magnitudes from the SFHs (SAM output), instead of during the SAM run time, where we adapt this part of Henriques et al. (2015) code to our lightcone construction script.

To be consistent with Millennium lightcones, we have used SED templates from Maraston (2005) stellar synthesis population models, assuming a Chabrier (2003) initial mass function. The templates contain 4×221 SEDs, with 4 different metallicities ($\log{(Z/Z_\odot)} = -1.35$, -0.33, 0.00, 0.35) and 221 different stellar ages (from ~ 0.2 Myr to ~ 20 Gyr). To attribute a SED to each SFH bin, we interpolate the four SEDs that are closest in age and metallicity, and multiply by the stellar mass of the bin.

The dust extinction was modelled following the approach used initially by De Lucia and Blaizot (2007) and, after, by Henriques et al. (2015), Shamshiri et al. (2015) and Clay et al. (2015). We present it here just for completeness. This dust model has two components: the interstellar medium (ISM) and molecular clouds (MC; actually birth molecular clouds) around

newly formed stars. The ISM extinction affects the light from disk stars, whereas the MC extinction acts on the light from young stellar populations, with ages $\leqslant 10$ Myr (Charlot and Fall, 2000). For the first component, the optical depth as a function of the wavelength is:

$$\tau_\lambda^{\text{ISM}} = \left(\frac{A_\lambda}{A_v}\right)_{Z_\odot} \left(\frac{Z_{\text{gas}}}{Z_\odot}\right)^s (1+z)^{-1} \times \left(\frac{\langle N_H \rangle}{2.1 \times 10^{21} \text{ atoms cm}^{-2}}\right) \qquad (2.9)$$

where $\langle N_H \rangle$ is the hydrogen mean column density. This is estimated from the SAM output parameters as

$$\langle N_H \rangle = \frac{M_{\text{cold}}}{1.4 m_p \pi (a R_{\text{gas},d})^2} \text{ atoms cm}^{-2} \qquad (2.10)$$

where M_{cold} is the mass of the cold gas, $R_{\text{gas},d}$ the gas disk radius, and $a = 1.68$. Using this value for a, $\langle N_H \rangle$ represents the mass-weighted average column density of an exponential disk. The factor 1.4 accounts for the helium abundance (Clay et al., 2015).

The $(Z_{\text{gas}}/Z_\odot)$ factor in Equation 2.10 represents the mass fraction of metals in the cold gas, in units of the solar metallicity, $Z_\odot = 0.02$. The s parameter depends on the wavelength: $s = 1.35$ for $\lambda < 2000$ Å, and $s = 1.60$ for $\lambda \geqslant 2000$ Å (Guiderdoni and Rocca-Volmerange, 1987). The extinction curve for solar metallicity, $(A_\lambda/A_v)_{Z_\odot}$, is extracted from Mathis et al. (1983).

The second extinction component, MC, affects only young stellar populations, as they are due to the remains of their progenitor molecular clouds. In this case,

$$\tau_\lambda^{\text{MC}} = \tau_V^{\text{ISM}} \left(\frac{1}{\mu} - 1\right) \left(\frac{\lambda}{5500 \text{ Å}}\right)^{-0.7} \qquad (2.11)$$

where τ_V^{ISM} represents the optical depth of the ISM in the V band ($\lambda_{eq} \sim 5500$ Å), and μ is a random Gaussian variable with values between 0.1 and 1, with mean 0.3 and standard deviation 0.2.

Therefore, we assume that the dust extinction of a galaxy can be written as

$$A_\lambda^{\text{ISM}} = \left(\frac{1 - e^{(-\tau_\lambda^{\text{ISM}} \sec \theta)}}{\tau_\lambda^{\text{ISM}} \sec \theta}\right)$$
$$A_\lambda^{\text{MC}} = (1 - e^{-\tau_\lambda^{\text{MC}}}) \qquad (2.12)$$

where θ represents the inclination of the galaxy. The inclination cosine is first randomly sampled between 0 and 1 and, after, all values smaller than 0.2 are set to 0.2 (Henriques et al., 2015).

Finally, the dusty SEDs of each galaxy are obtained by applying the ISM dust factor to the disk total luminosity and the MC factor to the luminosity of young stellar populations in the disk and/or bulge. Figure 2.6 presents a final SED of a galaxy at $z = 1.61$.
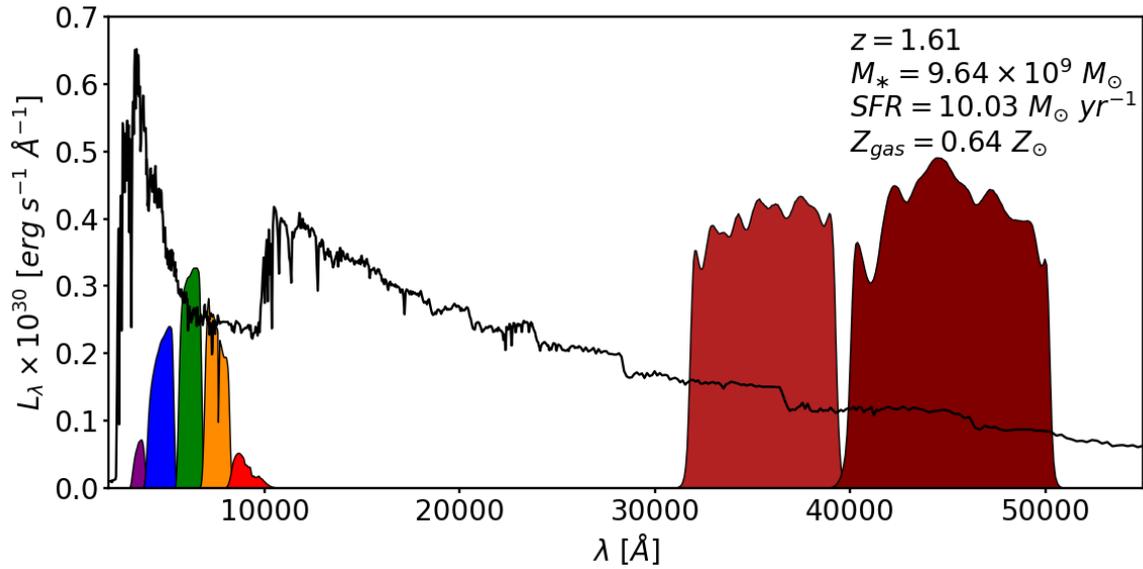
Figure 2.6: Dust corrected SED for a galaxy at $z = 1.61$, with $M_* = 9.64 \times 10^9\ M_\odot$, $SFR = 10.03\ M_\odot/\mathrm{yr}$ and gas metallicity $Z_{gas} = 0.64\ Z_\odot$. The filter transmissions are plotted just for illustration (u, g, r , i, z, IRAC3.6 and IRAC4.5, from left to right).

2.6 | Magnitude Estimations

As discussed in Blaizot et al. (2005), Kitzbichler and White (2007), and Merson et al. (2013), we cannot obtain the correct observer-frame magnitudes from the SAM output of the Millennium lightcones because the spectral deviation to be applied to estimate apparent magnitudes is the redshift of the snapshots from where the galaxies were extracted. Due to the discrete nature of these redshifts, the resulting magnitude and color distributions reflects this discreteness, mostly between the transition of two consecutive redshift snapshots (see Merson et al., 2013, for an example of this effect). To address this problem, Blaizot et al. (2005) has developed the Mock Map Facility (MoMaF), where an interpolation is applied to these discrete magnitudes, according to the position of the galaxies within the comoving box and their velocities, in order to obtain a better approximation for the observer-frame magnitudes. This tool is available in the L-GALAXIES SAM, but, since the origin of the Millennium volume in the lightcone space is different from that of our work, this approach is not directly applicable to our lightcones. It can be shown, however, as discussed by Shamshiri et al. (2015), that the root-mean-square difference between the magnitudes that are computed during the SAM run time and the magnitudes obtained as *post-processing* do not exceed 0.12 mag for the u band (the worst case for optical bands) for $z = 2.0$ galaxies and that the difference decreases for redder filters (0.02 mag in the IRAC-4.5μm filter) and lower redshifts. It is also important to remark that the magnitudes estimated during the SAM run time have more time resolution than *post-processing* magnitudes. The former is

computed, tracking the star formation from 1160 timesteps, for $z = 0$ galaxies, while the number of SFH bins are limited to 20.

Since we have attributed a SED to each mock galaxy, we can directly obtain apparent magnitudes using the considering the filter response function and the galaxy redshift estimated as discussed in Section 2.4. The observer-frame flux S_ν is given by

$$S_\nu = (1 + z)\frac{L_{(1+z)\nu}}{4\pi d_L(z)^2} \tag{2.13}$$

where $L_{(1+z)\nu}$ is the luminosity at the frequency $(1+z)\nu$ for a galaxy at redshift z and luminosity distance $d_L(z)$.

We estimate apparent magnitudes in the AB system (Oke and Gunn, 1983) as

$$m_{\mathrm{AB}} = -2.5\log_{10}\left[\frac{\int S_\nu R(\nu)d\nu}{S_{\mathrm{o}}\int R(\nu)d\nu}\right] \tag{2.14}$$

where $R(\nu)$ is the filter transmission and S_{o} is the zero-point of the AB system, $S_{\mathrm{o}} = 3631$ Jy.

We compute magnitudes for 31 broad bands from FUV to MIR. Despite the filter transmission of HSC and LSST being similar, we use the filter transmission curves corresponding to each instrument.

2.7 | IGM Absorption

Magnitudes and colors of high redshift galaxies are critically affected by the absorption of rest-frame UV photons in the intergalactic medium (IGM), mostly those that go through neutral hydrogen clouds, optically thin systems, and resonant scattering of Lyman transitions. Photons with a wavelength shorter than 1216 Å (Lyα) are easily absorbed by neutral hydrogen because they excite or ionize HI atoms, producing a deficit on the observed flux that is more critical for energies higher than those of the Lyman Limit (912Å). In consequence, we cannot detect, in the observer-frame, certain objects in the bluer filters. For example, a $z = 3.0$ galaxy will be an u-dropout.

In order to construct a more realistic mock catalog, we adopt the same method as Overzier et al. (2013) for IGM absorption correction, which we summarize here for completeness. Overzier et al. (2013) correct the mock magnitudes after the filter convolution, as *post-processing*. `L-GALAXIES` does not bring information on the IGM gas density. However, this quantity can be statistically addressed by an IGM correction based on Monte Carlo simulations, where an effective optical depth is computed from the sum of the contributions of the attenuation sources, using the Inoue and Iwata (2008) IGM model. Since the number of absorbers along a certain line-of-sight depends on the redshift, attenuation curves are computed in redshift bins spaced by 0.1, using the `IGMtransmission` code (Harrison et al., 2011). This code performs Monte Carlo simulations

for 10,000 different line-of-sights and then estimate the effective optical depth. The attenuation curves are applied on a 100 Myr old, continuously star-forming, solar metallicity SED, modeled using `starburst99` (Leitherer and Heckman, 1995). After, the differences between the IGM corrected and the intrinsic magnitudes in all photometric filters are determined.

2.8 | Protocluster-Lightcones

A particularity of our lightcones is that we can place desired structures at specific redshifts.

We have chosen randomly 20 Millennium $z = 0$ galaxy clusters, and considered the progenitor of these 20 clusters at $z = 1.0$, 1.5, 2.0, 2.5, and 3.0. Following the definition of Chiang et al. (2013), we have 8 Fornax-type ($M_{z=0} = 1.37 - 3.00 \times 10^{14}$ $M_\odot$), 6 Virgo-type ($M_{z=0} = 3 - 10 \times 10^{14}$ $M_\odot$) and 6 Coma-type ($M_{z=0} \geqslant 10^{15}$ $M_\odot$) protoclusters. We present in Figure 2.7 the "observational" 3-D coordinates of 3 of these protoclusters at $z = 1.0$, 2.0, and 3.0, where we highlight the galaxy members of these structures. We select the protocluster members as all galaxies that reside in the dark matter halos that will evolve into the chosen cluster. We present in Table 2.1 some observational properties of these protoclusters. We estimate the median cluster galaxy angular distance to the cluster centroid ($\delta\theta_p$), the velocity dispersion ($c\sigma_z/(1+z)$ where σ_z is the standard deviation of the radial velocities of cluster galaxies), and the number of members with $M_\star \geqslant 10^8$ $M_\odot/h$ (N_{gal}). Since these quantities depend on the evolutionary state of the structure, we present them at the five different redshifts.

This type of lightcone is useful to increase the number of rare structures at specific redshifts, in particular the progenitors of massive clusters (Coma-type). Besides, they can help to track the evolution of structures from an observational point of view, as shown in Table 2.1.

Table 2.1: Descendant mass, median separation between members and the centroid, velocity dispersion, and number of protocluster members the 20 placed structures at $z = 1.0$, 1.5, 2.0, 2.5 and 3.0.

$\log\left(M_{z=0}/M_\odot\right)$	$\delta\theta_p$ [arcmin]					$c\sigma/(1+z)$ [km/s]					N_gal				
	$z=1.0$	$z=1.5$	$z=2.0$	$z=2.5$	$z=3.0$	$z=1.0$	$z=1.5$	$z=2.0$	$z=2.5$	$z=3.0$	$z=1.0$	$z=1.5$	$z=2.0$	$z=2.5$	$z=3.0$
14.14	0.80	1.40	1.62	1.77	1.80	257.79	254.71	231.02	141.98	165.27	74	74	76	80	78
14.17	0.80	1.25	1.47	1.59	1.59	467.73	278.84	199.33	193.88	141.79	97	101	93	90	88
14.20	0.78	0.92	1.03	1.19	1.22	380.46	245.03	322.52	205.25	217.87	75	80	95	94	85
14.23	0.76	1.08	1.17	1.11	1.06	398.82	260.19	212.49	160.65	128.92	74	71	77	71	74
14.27	1.89	2.22	2.49	2.34	2.18	276.48	206.40	165.45	140.33	127.16	81	91	95	95	82
14.31	1.15	1.78	1.55	1.77	1.71	364.60	269.56	199.26	194.50	165.96	90	111	111	111	106
14.36	0.83	1.06	1.34	1.53	1.74	307.77	282.04	280.35	210.81	184.77	54	59	57	60	56
14.42	1.52	1.60	2.19	2.24	2.24	416.96	362.18	284.29	233.02	177.78	125	137	141	127	114
14.49	0.92	1.09	1.13	1.25	1.44	407.09	396.81	281.18	243.42	242.17	109	129	119	131	116
14.58	1.41	1.21	1.37	1.38	1.62	376.47	325.81	243.41	225.44	212.62	109	116	122	115	97
14.73	0.89	1.18	1.41	1.64	1.77	647.35	500.49	443.68	325.70	297.03	209	251	262	269	263
14.73	1.80	3.22	3.11	3.05	2.82	421.83	270.74	247.66	210.62	223.00	205	235	239	218	198
14.82	1.36	1.64	1.83	1.98	1.89	527.48	377.53	296.57	316.65	356.97	296	330	322	344	321
14.94	1.62	2.39	2.42	2.68	2.66	626.69	489.45	415.85	282.60	249.00	385	443	438	423	402
15.03	2.69	2.52	2.56	2.61	2.62	542.45	398.31	326.22	334.36	335.15	352	369	365	354	344
15.05	0.92	1.39	1.95	2.29	2.38	918.87	589.27	365.95	273.45	258.58	432	466	462	443	431
15.07	3.18	3.85	4.05	4.04	3.81	548.90	425.99	396.73	370.91	324.81	402	426	423	411	389
15.16	4.16	3.76	3.45	3.47	3.35	508.83	450.26	339.91	341.90	320.92	451	448	442	437	414
15.26	1.98	2.12	1.91	2.08	1.97	452.64	299.46	273.02	288.02	266.29	300	292	291	272	266
15.36	2.78	3.56	3.47	3.54	3.56	651.56	481.89	427.88	457.91	504.98	901	973	960	918	839
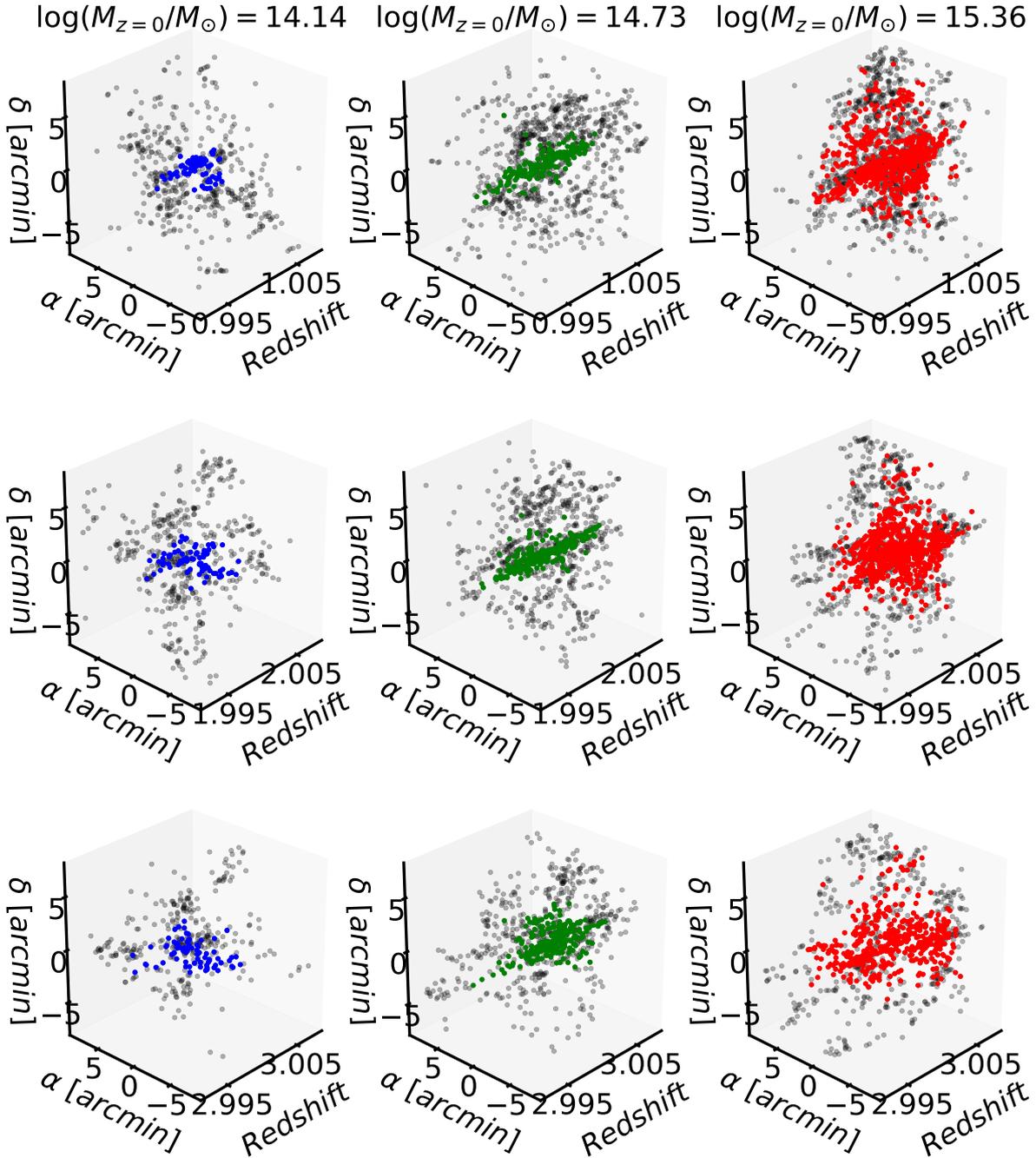
Figure 2.7: 3-D spatial celestial coordinate distribution (right ascension (α), declination (δ) and redshift) of three placed protoclusters at $z = 1.0$, 2.0 and 3.0. Colored dots represents the protocluster members.

2.9 | Validation Tests

To test the agreement of the PCcones with observations and other publicly available mock catalogs, we now compare magnitude and color distributions as a function of the redshift. We also compared our results with those obtained from the lightcone constructed with the Henriques et al. (2015) SAM, which uses the `MoMaF` code to obtain observer-frame magnitudes by interpolation.
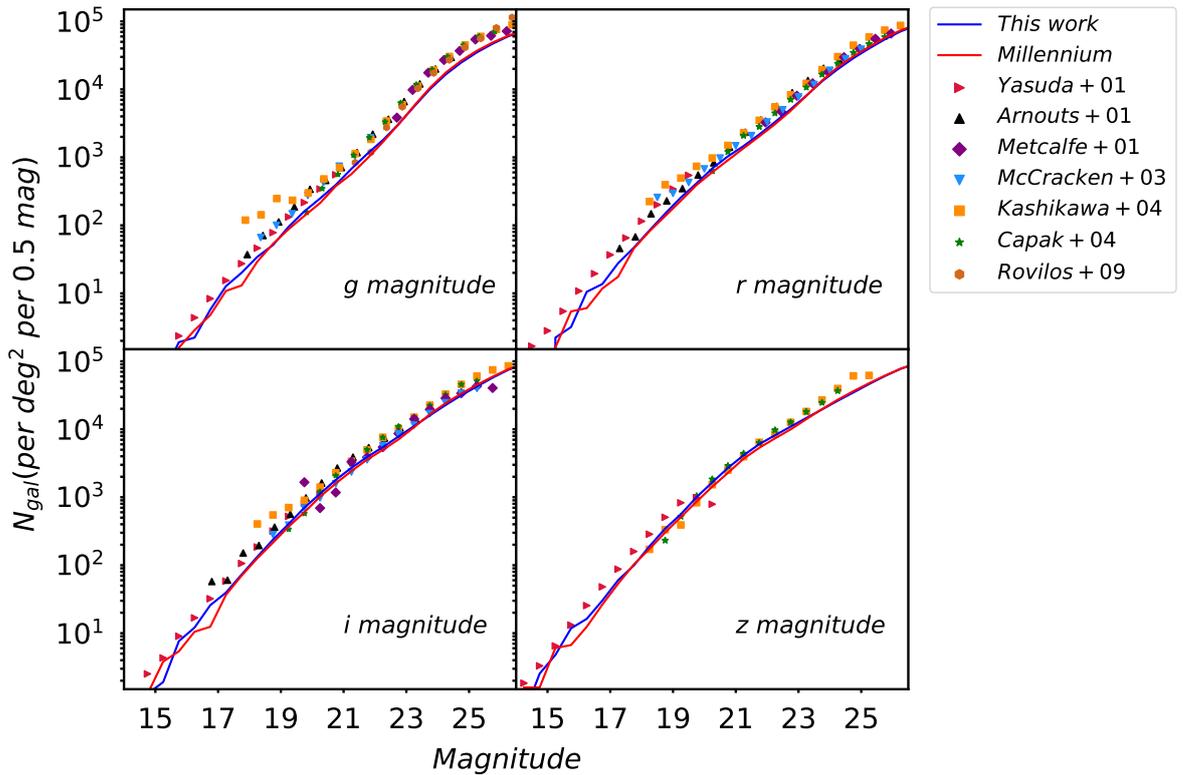
Figure 2.8: Number of galaxies per magnitude bin of width 0.5 per 1 deg^2 of four Sloan filters: g (upper left), r (upper right), i (bottom left) and z (bottom right) bands. Both simulated magnitude distributions (blue solid and red dashed lines for this work and the Millennium lightcone, respectively) are compared with observational data (other symbols) from: Yasuda et al. (2001), Arnouts et al. (2001), Metcalfe et al. (2001), McCracken et al. (2003), Kashikawa et al. (2004), Capak et al. (2004) and Rovilos et al. (2009).

2.9.1 | Galaxy Counts

Figure 2.8 compares galaxy differential number counts as a function of the magnitude in several bands, with observational data from a variety of sources[2]. We compare the predicted magnitude distributions with data from Yasuda et al. (2001), Kashikawa et al. (2004) and Capak

---

[2]data from `http://astro.dur.ac.uk/~nm/pubhtml/counts/`

et al. (2004) for the four photometric bands, while for g, r and i filters, we also show Arnouts
et al. (2001); Metcalfe et al. (2001) and McCracken et al. (2003) data. Rovilos et al. (2009)
galaxy counts are presented for the g band.

Galaxies with stellar masses higher than 10^8 $M_\odot/h$ were considered to construct both pre-
dicted galaxy counts (this work and Millennium Lightcone) and, as expected, both are consistent
with each other since they were computed using the same simulation. Magnitude distributions
show consistency with observational data in all analyzed bands at magnitudes brighter than
$g \sim 27.0$ mag.

2.9.2 | Color-Redshift Distributions
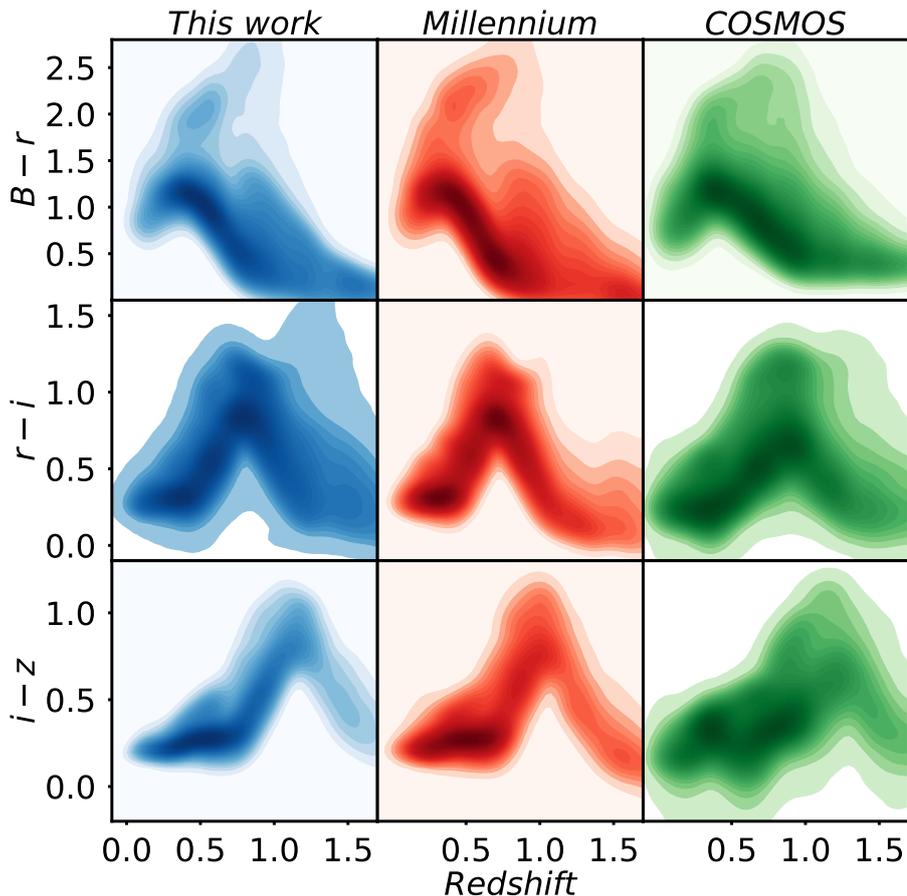
Figure 2.9: Color as a function of redshift for our lightcones (first column), Millennium lightcones (second
column), and for the COSMOS data (third column). Each row represents a different color: $B - r$ (first
row), $r - i$ (second row) and $i - z$ (third row).

We show, in Figure 2.9, the $B - r$, $r - i$ and $i - z$ colors as a function of the redshift for our
mocks, Millennium simulation and for the COSMOS data. The two main galaxy populations, red
and blue galaxies, can be seen in these color-redshift distributions, both in predicted distributions

(this work and the Millennium lightcones) as well as in those obtained from COSMOS data. Indeed, both Figures 2.8 and 2.9 present good qualitative agreement between color distributions in the observational and simulated data. This is assuring, considering that a major motivation for numerical simulations is to produce as realistic as possible mock catalogues to forecast galaxy survey observations.

3 Photometric Redshifts

The redshift of the galaxies is related to their distance to us, according to the cosmological model, as used in section 2.4. This parameter allow us to obtain the 3-D spatial distribution of the galaxies in the universe, which is required to detect real structures from observations. Redshifts are measured directly from the galaxy spectra by identifying the shift of the most prominent emission or absorption lines. In general, we need long exposure times to obtain this information, and even more for distant sources. Then, spectroscopic data of the high redshift universe is still uncomplete because their obtention is very challenging. An alternative are the photometric redshifts (photo-zs), whose estimation is based on galaxy magnitudes and colors. In general, the accuracy of these measurements ranges from $\sigma_z \sim 0.02 - 0.06$, depending on the number of bands and photometric errors.

There are two main methods to estimate photo-z: with template fitting algorithms or machine learning techniques. The former uses theoretical or empirical SEDs of different types of galaxies (spirals, ellipticals, lenticulars, irregulars), which have to represent the galaxy populations in the universe. From them, observer-frame magnitudes are computed, applying the spectral deviation effect at different redshifts. After, theoretical (from templates) and observed magnitudes are compared with a χ^2 minimization method, obtaining the most likely redshift of the galaxy. Some examples of SED fitting codes include BPZ (Benítez, 2000), Le-Phare (Arnouts et al., 2002; Ilbert et al., 2006), EAZY (Brammer et al., 2008), among others.

The second main method to estimate photometric redshifts is with machine-learning. Some codes, such as ANNz (Sadeh et al., 2016), GPZ (Almosallam et al., 2016), among others, include different techniques to perform regressions. Examples include Support Vector Machine, Decision Trees, Nearest Neighbors, Gaussian Processes, and others. Also, we can create our own Neural Network, using the Keras Python Package, allowing us implement deep-learning algorithms without extensive programming efforts.

The most important issue of this method is that the training dataset has to present consistency with the data for which we will compute photo-zs. The adopted algorithm is trained with input features, mostly colors, and magnitude in a single band, from a dataset with spectroscopic redshift. Generally, this approach achieves higher accuracy than SED fitting (Lima et al., in prep.), when a complete training set is available. However, for high redshift and faint objects, the spectroscopic samples are still uncomplete and inhomogeneous. Thus, they are still not reliable for application in new deep surveys.

Photometric redshifts are often adopted as a tool to identify structures in the universe, either from targeted observations (e.g. Overzier et al., 2008; Hatch et al., 2017; Watson et al., 2019; Strazzullo et al., 2019), or in photometric surveys (e.g. O'Mill et al., 2011; Sánchez et al., 2014; Bilicki et al., 2018; Molino et al., 2019a). Photometric redshifts are also useful to find galaxy overdensities, and mock catalogues can then be used to quantify the uncertainties of this

approach. Chiang et al. (2013) have examined the impact on simulated protocluster overdensities due to redshift uncertainties, showing that, depending on the galaxy tracer population used, the size of the region, and the redshift accuracy, random regions can present similar overdensities as cluster progenitors at $z = 3$. The back- and foreground contamination is often large in photometrically selected overdense regions, as can be seen with a spectroscopic follow-up (e.g. Dey et al., 2016).

We now readdress the impact on overdensity estimations of uncertainties in photometric redshifts with our PCcones. For this, we will apply to the mocks the same constraints of observational surveys. We will emulate three optical photometric surveys: the Deep Canada-France Hawaii Telescope Legacy Survey (CFHTLS; https://www.cfht.hawaii.edu/Science/CFHTLS/), which observed 3.2 deg^2 of the sky in 5 photometric bands; the ongoing, Hyper Supreme Cam Subaru Strategic Program (HSC-SSP) (Aihara et al., 2018), whose wide-layer will cover 1,400 deg^2, where more than 300 deg^2 were already observed; and, finally, the future Vera Rubin Observatory Legacy Survey of Space and Time (LSST) (Ivezic et al., 2008), which will provide deep photometric information in 6 bands over ∼20,000 deg^2 of the southern sky after 10-year of operation.

We adopt for the photometric redshift estimation the `Le Phare` software (Arnouts et al., 2002; Ilbert et al., 2006), using the same set of SEDs presented by Ilbert et al. (2009), also adopted by Ilbert et al. (2013) and Laigle et al. (2016). They consist of 31 templates, including spiral and elliptical galaxies from Polletta et al. (2007) (a total of 19) and also 12 young blue star-forming galaxies modeled by using Bruzual and Charlot (2003) stellar population SEDs. We added dust extinction as a free parameter ($E(B-V) < 0.5$), and considered different extinction laws: Calzetti et al. (2000), Prevot et al. (1984), and the Calzetti laws including a bump at 2175 Å (Fitzpatrick and Massa, 1986).

3.1 | Observational-like magnitudes

Observed galaxy magnitudes are affected by many factors, such as the exposure time, sky brightness, quantum efficiency of the detector, the point spread function etc, that can introduce systematic and random errors in the measurements. In order to emulate observational magnitudes in our mocks, we have implemented the same technique as Graham et al. (2018) where we assume random errors in the mock magnitudes by using the analytic expression presented in Ivezić et al. (2019):

$$\sigma_{rand}^2 = (0.04 - \gamma)x + \gamma x^2 \ (mag^2), \tag{3.1}$$

where $x \equiv 10^{0.4(m-m_5)}$, m_5 corresponds to the 5σ magnitude limit of the observational survey to be emulated, and γ depends on sky brightness, signal-to-noise and photometric filter. We set γ to the same values presented in Table 2 of Ivezić et al. (2019) for optical magnitudes. The

values of m_5 and γ that we adopt to simulate the deep survey of CFHTLS, the wide survey of HSC-SSP and LSST, are listed in Table 3.1, where we adopt a magnitude dependent γ that is the same for the three surveys.

Table 3.1: List of values used for observational like magnitudes estimations for ground based surveys (the deep survey of CFHTLS, the wide survey of HSC-SSP and LSST) for each filter, where m_5 represents the 5σ magnitude limit and γ a image quality parameter.

Filter	m_5 (CFHT)	m_5 (HSC)	m_5 (LSST)	γ
u	26.3	-	26.1	0.038
g	26.0	26.8	27.4	0.039
r	25.6	26.4	27.5	0.039
i	25.4	26.2	27.0	0.039
z	25.0	25.4	26.1	0.039
y	-	24.7	24.9	0.039

The total magnitude error is obtained as: $\sigma_{mag} = \sqrt{\sigma_{sys}^2 + \sigma_{rand}^2}$, where we have assumed a systematic error of $\sigma_{sys} = 0.005$, following Ivezić et al. (2019).

3.1.1 | Non-detected sources

As we have mentioned in Section 2.7, the Lyman Break may lead high redshift galaxies not being detected in all photometric filters due to IGM absorption along the line of sight. Also, the depth of each survey plays an important role in the sources that we can observe. Since mock catalogs are limited just by the resolution of the simulation, and this limit allows us to have a complete sample at fainter magnitudes compared with the limits of the surveys discussed here, we need to allow for non-detected sources in some photometric bands. Therefore, we define as a non-detected source all galaxies with signal to noise smaller or equal than one for each of the three surveys, and we set this value as an upper limit in the photometric redshift estimation with Le Phare.
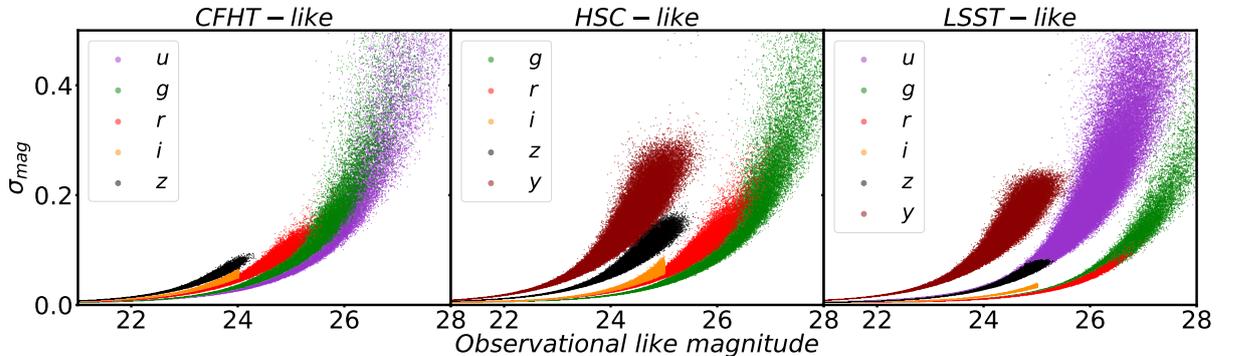
Figure 3.1: Magnitude error as a function of observational-like magnitudes for deep CFHTLS (first panel), the wide survey of HSC-SSP (second panel), and LSST (third panel). Observational-like magnitudes are obtained from Equation 3.1 using parameters listed on Table 3.1 for CFHTLS, HSC-SSP, and LSST like magnitudes.

3.2 | Comparison with Millennium photometric redshifts

In this section we compare photometric redshift estimates obtained for the PCcones and Millennium lightcones. We simulate observations of the deep Canada France Hawaii Telescope Legacy survey using both lightcones. The values of m_5 and γ for each filter are listed on Table 3.1. The photo-z estimation performed with `Le Phare` are presented in Figure 3.2.

To quantify the accuracy of these photometric redshifts we estimate the normalized median absolute deviation, σ_{NMAD}, defined as in Molino et al. (2019b). Also, we quantify the *bias*, and the outlier fraction, f_{outliers}, following Tanaka et al. (2018). We calculate these quantities as

$$\sigma_{\mathrm{NMAD}} = 1.48 \times median\left(\frac{|\delta z - median(\delta z)|}{1 + z_r}\right)$$
$$f_{\mathrm{outliers}} = \frac{N[|\delta_z|/(1 + z_r) > 0.15]}{N_{total}} \tag{3.2}$$
$$bias = \langle \delta z \rangle$$

where $\delta z = z_p - z_r$, with z_p denoting the estimated photometric redshift and z_r the reference redshift of the galaxies in the lightcone. We obtain $\sigma_{\mathrm{NMAD}} = 0.0440$, $bias = 0.042$ and $f_{\mathrm{outliers}} = 0.143$ for the Millennium lightcone, while for the photo-z estimation over our lightcones we achieve $\sigma_{\mathrm{NMAD}} = 0.0275$, $bias = -0.004$ and $f_{\mathrm{outliers}} = 0.034$.

These results and Figure 3.2 (left panel) indicate that photometric redshifts using our lightcones are more reliable than those with the usual Millennium lightcones. Compared to ours, the photo-z estimation using the Millennium apparent magnitudes presents a clear bias for $z \lesssim 1$, where the linear correlation is offset with respect to the equal values line. Additionally, the bias and the outlier fraction obtained with the conventional Millennium simulation are larger than ours.

The unique main difference between both lightcones is the method to estimate apparent magnitudes, in particular, the observer-frame magnitudes. We computed magnitudes in *post-processing* mode, following Shamshiri et al. (2015), attributing a SED to each galaxy, while those of the Millennium Lightcones come from interpolations using MoMaF (Blaizot et al., 2005). To confirm that the differences in photo-z estimates are indeed caused by this, we have performed a more detailed comparison. We re-calculated the apparent magnitudes for the galaxies in the Millennium Lightcone at $z \leqslant 3.0$ with our method. Magnitude distributions are similar to each other, as Figure 2.8 shows. However, we found that the median of the absolute difference between galaxy magnitudes, $median(|m_{\mathrm{our}} - m_{\mathrm{ml}}|)$ are ~ 0.23, 0.18, 0.19, 0.21 and 0.24 mag for the u, g, r, i and z photometric bands, respectively.

After, we generated observational-like catalogs, emulating again the Deep CFHTLS, and then we run `Le Phare`. Since we obtain observational-like magnitudes as Gaussian random values with median mock magnitudes and standard deviations equal to the expected errors, the galaxies are

not anymore the same in both catalogs. However, we find that $\sim 86\%$ of the galaxies are in both observational-like catalogs. The right panels of Figure 3.2 show galaxies that were found in both datasets. This figure is very similar to the left panels of Figure 3.2 and indicates that the difference between the quality of our estimated photometric redshifts and those obtained from the Millennium Lightcone is due to the different approaches in estimating the magnitudes of each catalogue. Also, we obtain similar values of σ_{NMAD}, bias, and outlier fraction.

Finally, we conclude that photometric redshifts using our lightcones are more reliable because our method to obtain observer-frame magnitudes implicitly change the shape of the galaxy SED caused by spectral deviation, while the interpolated ones are only corrected by the systematic wavelength shift (Merson et al., 2013). In addition, `Le Phare` fits real SEDs, therefore, the differences in the photo-zs estimates could come from the fitting of a real SEDs to an interpolated color.
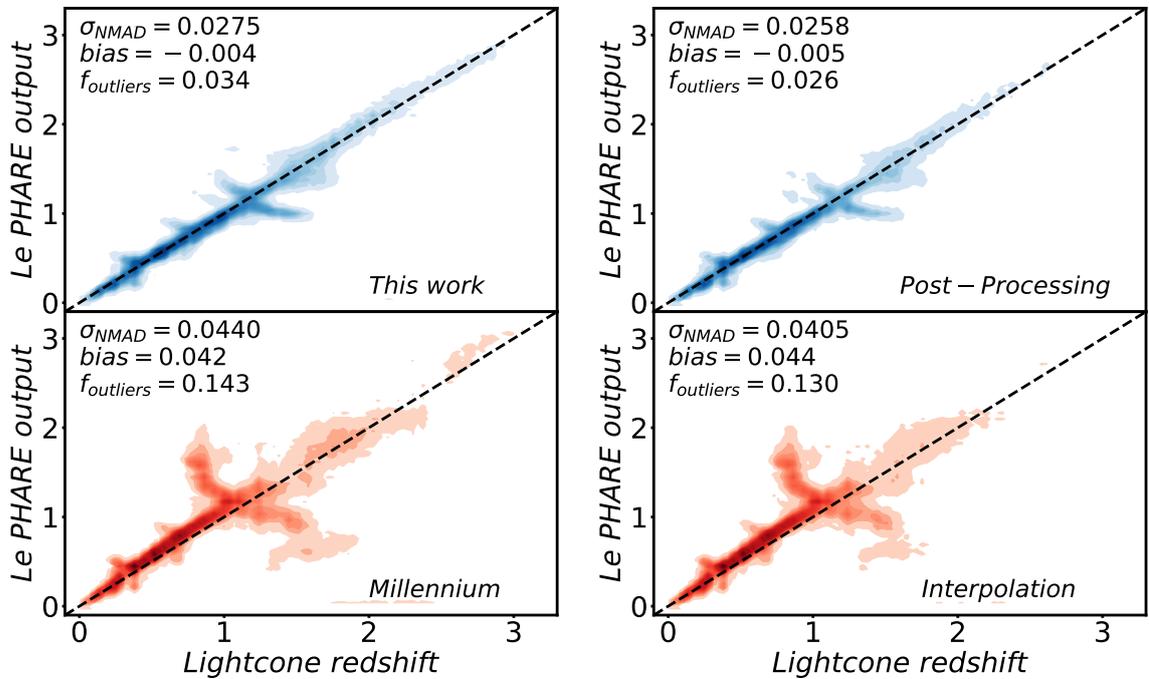
Figure 3.2: Left: Photometric redshift estimation using `Le Phare` versus the lightcone redshift for this work (blue contours, first panel) and the Millennium Lightcone (red contours, second panel). Photo-zs were estimated from the photometric information of the deep CFHTLS. Right: Photometric redshift estimations. We compare results from the Millennium Lightcone using *post-processing* apparent magnitudes (blue contours, first panel) with those obtained with interpolated magnitudes (red contours, second panel). Photo-zs were estimated with the photometric constraints of the deep CFHTLS.

3.3 | Photometric Redshifts for Simulated Surveys
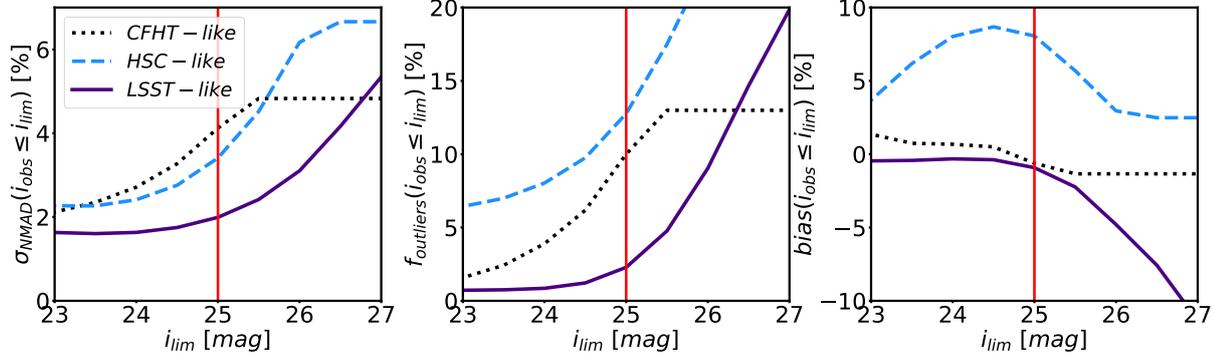
Figure 3.3: Distribution of the normalized median absolute deviation (σ_{NMAD}, first panel), outlier fraction (f_{outlier}, second panel), and *bias* (third panel) as a function of magnitudes for the CFHT-, HSC-, and LSST-like emulations. Red vertical line indicates the imposed magnitude limit for the full mock sample.

We now use our lightcones to emulate photometric redshift estimation for some actual photometric surveys. To mimic observations of the deep CFHT legacy survey, the wide layer of the Hyper Suprime Cam Strategic Survey Program (HSC-SSP), and LSST, we assume the 5σ magnitude limits listed in Table 3.1 in the photometric bands available for each of these surveys. Firstly, we impose a magnitude cut in the $i-$band equals to the 5σ limit to analyze the photo-z accuracy according to our three metrics: σ_{NMAD}, f_{outlier} and *bias* (see Equation 3.2) for just one mock. Figure 3.3 show the dependence of these parameters as a function of the magnitude in the $i-$band. This figure recovers an important point in photometric redshift estimation, namely, the improvement in photo-z accuracy (σ_{NMAD}) by increasing the number of bands and by decreasing photometric errors. Additionally, all the analyzed parameters reach catastrophic values for fainter magnitudes. For example, the *bias* parameter starts to fall quickly for LSST at $i > 25.0$ mag, while the outlier fraction and σ_{NMAD} grow exponentially for HSC-SSP and LSST at the same range. For this reason, we have estimated photometric redshifts for the whole PCcone sample, but constrained to objects brighter than $i = 25.0$ mag for HSC-SSP and LSST mock surveys. In the case of the CFHTLS emulation, we imposed a magnitude cut at $i = 24.0$ mag, the same as Wen and Han (2011), who produced a cluster sample that we will emulate here (see Section 4.2). Figure 3.4 shows our photometric redshift estimates against the galaxy redshift in the lightcones for the applied magnitude limits.

We have obtained $\sigma_{\mathrm{NMAD}} = 0.027$, 0.033, and 0.020 for CFHTLS, HSC-SSP and LSST-like observations, respectively. Since our definition of accuracy differs from those reported in papers related to these surveys, below we use the same metrics adopted in these other studies, the normalized redshift dispersion. Ilbert et al. (2006) obtained a redshift accuracy of about $\sigma_z/(1+z) = 0.029$ for the deep CFHTLS, for galaxies with $i < 24.0$ mag, while we obtained $\sigma_z/(1+z) = 0.027$. Also, Graham et al. (2018) predicted a photo-z accuracy about $\sigma_z/(1+z) = 0.0165$ for the 10 years of LSST for galaxies with $i < 25.0$ mag. After applying this same

magnitude limit, however, we achieve $\sigma_z/(1+z) = 0.020$. For the HSC-like sample, we obtained $\sigma_z/(1+z) = 0.037$, while Tanaka et al. (2018) obtained an accuracy of about $\sigma_z/(1+z) = 0.05$.

For the outlier fraction, $f_{outlier}$, Ilbert et al. (2006) obtained for the Deep CFHTL $f_{outlier} = 3.8\%$ at $i < 24.0$ mag, while we achieved a $f_{outlier} = 3.9\%$ with the same magnitude restriction. For the wide-layer of HSC-SSP, Tanaka et al. (2018) estimated that a ~ 15 % of the galaxies present catastrophic redshifts, whereas our measurements indicate a $\sim 13\%$. Finally, for the LSST case, Graham et al. (2018) computed this fraction being more strict than the previously mentioned reports. They considered as outliers all estimates than differ by a factor 0.06 instead 0.15 in the second row of Equation 3.2. Using the same definition of Graham et al. (2018), that found a fraction of 4% of outliers, we predict for the 10-year forecast of LSST a fraction of $\sim 3\%$. Notice that for HSC-SSP and LSST, the applied magnitude limit in this work and the others is $i = 25.0$ mag.
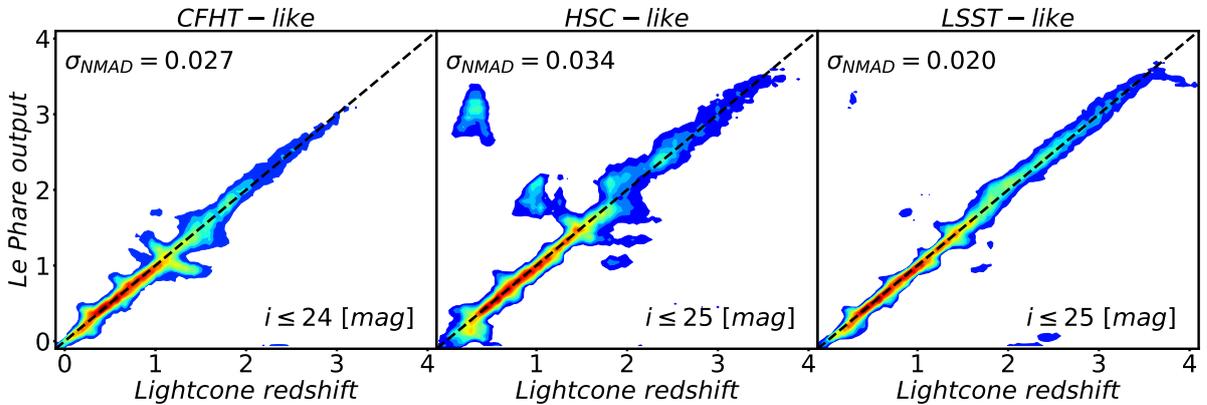
Figure 3.4: Photometric redshift estimation using `Le Phare` for the deep CFHTLS (left panel), the wide layer of HSC-SSP (middle panel), and LSST (right panel) survey emulations. These results show that our procedure gives photometric redshifts with accuracy comparable to those inferred from the studies mentioned above.

4 | Results

Measuring the excess of galaxies in a given region of the universe with respect to the mean distribution is one of the most common techniques to look for cluster/protocluster candidates. In this section we explore overdensity measurements taking in to account observational constraints on our mocks.

We shall divide our discussion into two parts: in the first one, we will perform a forecast of structure detection in specific redshift ranges ($z = 1.0, 1.5, 2.0, 2.5$ and 3.0), emulating the wide layer of the HSC-SSP and the future LSST survey. We compare these mock overdensity distributions, measured in redshift intervals (see below), to those obtained from mocks without magnitude constraints. Also, we estimate the probability of a region with overdensity δ_{gal} be a protocluster.

In the second sub-section, we compare the cluster distribution for a CFHTLS mock with actual observational data in the redshift range $0.12 < z < 1.70$, as an additional test of using our lightcones to predict cluster detections over a wide redshift range.

4.1 | Structure Detection Forecast

In this section we describe our method to detect protoclusters by the construction of density contrast maps at different redshifts. We use this information to compute the probabilities of a certain region be a real structure. Also, we analyze the detectability of the protoclusters placed in the middle of the line-of-sight in the PCcones.

4.1.1 | Mock samples

We perform our analysis for two main different mock samples: the first one (which we call *Pure Simulation*) contains all simulated galaxies (i.e., without any magnitude constraints besides those intrinsic to the Millenium simulation) within a Δz according to the photometric redshift accuracy achieved for each survey, centred at the redshift where we placed the protocluster. Our motivation with this sample is to compare overdensity measurements in mocks without any

additional magnitude limits and using redshift from simulations, to estimations from the photo-z surveys emulations. We estimate *Pure Simulation* overdensities within $\Delta z = \sigma_{\mathrm{NMAD}}^{HSC} \times (1+z) = 0.034 \times (1+z)$ and $\Delta z = \sigma_{\mathrm{NMAD}}^{LSST} \times (1+z) = 0.020 \times (1+z)$ for HSC and LSST, respectively.

The second type of samples, which we call *Observational-like* samples, is composed by the two mock observational surveys, emulating observations of the HSC-SSP wide layer (hereafter, *HSC-like*) and the 10-year LSST survey forecast (hereafter, *LSST-like*). We have done this by adding noise to the mock magnitudes. As explained in Section 3.1, this depends on the 5σ magnitude limits of the surveys, listed in Table 3.1. For both surveys, we considered all galaxies with $i \leqslant 25.0$ mag. We also use the photometric redshifts we have estimated before, which have a mean accuracy of $\sigma_{\mathrm{NMAD}} = 0.034$ and 0.020 for the *HSC-like* and *LSST-like* samples, respectively, up to the applied magnitude limit.

4.1.2 | Overdensity Estimation

We estimate overdensities in redshift slabs by selecting galaxies in $z_p \pm \Delta z$, where z_p is the analyzed redshift (1.0, 1.5, 2.0, 2.5, or 3.0) and $\Delta z = \sigma_{\mathrm{NMAD}} \times (1+z)$, similar to Chiang et al. (2014). These authors adopted $\Delta z = 0.0125 \times (1+z)$, according to the photometric redshift accuracy of the COSMOS/ultraVISTA survey. Within each slab, the density field is computed using the Gaussian Kernel Density Estimator (KDE) of the `scipy` Python package, version 1.3.2. The kernel bandwidth is set as a function of redshift as $\delta\theta(z) = R_e(z)$, where $R_e(z)$, following Chiang et al. (2013), is the effective radius of Coma-type protoclusters. It takes values: $R_e(z) = 3.5, 5.5, 6.5, 7.5$ and 8.2 comoving-Mpc at $z = 1.0, 1.5, 2.0, 2.5$ and 3.0, respectively.

If Σ is the Gaussian kernel estimator surface density for a given celestial coordinate in a redshift slab, the density contrast, δ_{gal}, at this point is given by:

$$\delta_{\mathrm{gal}} = \frac{\Sigma - \langle\Sigma\rangle}{\langle\Sigma\rangle} \tag{4.1}$$

where $\langle\Sigma\rangle$ denotes the mean of Σ inside the slab. Examples of surface density maps are presented in Figure 4.1 for the two mocks discussed in this section, where the progenitor of a $z = 0$ galaxy cluster of $M_{z=0} = 2.28 \times 10^{15} M_\odot$ has been placed at redshift $z = 1.0$. The density contrast maps are computed using a 200×200 pixels grid that covers $-1.0 < \Delta\, RA < 1.0$ [deg] and $-1.0 < \Delta\, Dec < 1.0$ [deg]
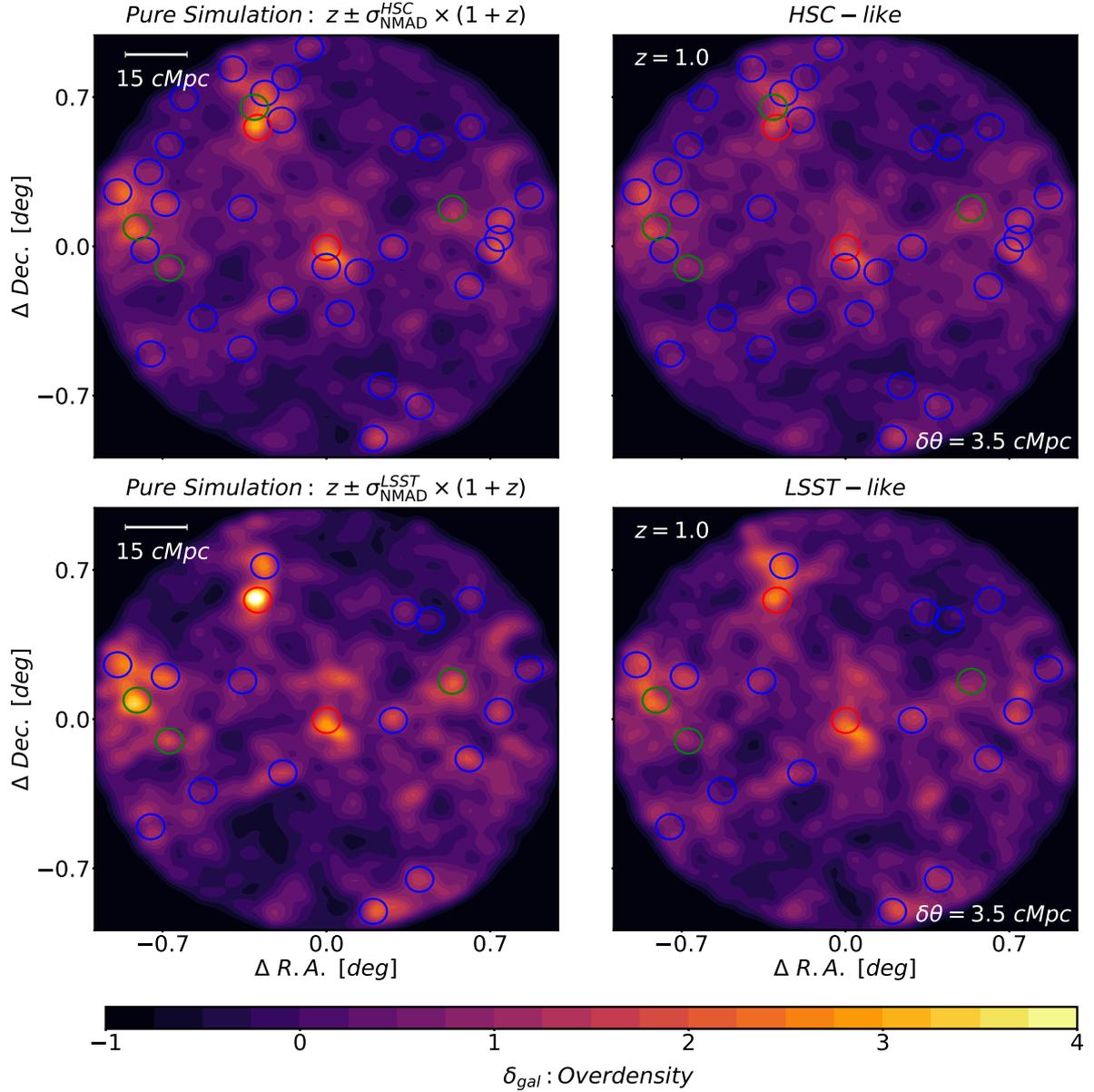
Figure 4.1: Example of the surface density map obtained from a PCcone where a $\log\left(M_{z=0}/M_\odot\right) = 15.26$ cluster progenitor is placed at the center of the lightcone at redshift $z_p = 1.0$. We present the density map for the *Pure Simulation* sample (left column) and the *Observational-like* samples (right column). HSC-like simulations are at the top, whereas LSST-like ones are at the bottom both assuming a magnitude limit $i \leqslant 25.0$ mag. The color circles, with radius R_e, mark protocluster regions, where the red, green and blue ones indicate Coma, Virgo and Fornax type protoclusters, respectively . Notice that there are several other protoclusters in the slab, besides that put at its centre.

4.1.3 | Pure Simulation Overdensities

In this section, we analyze the *Pure Simulation* sample, which considers all simulated galaxies in a certain volume, without any photometric constraints. We look for the (proto)clusters in redshift slabs within the lightcone field of view.

In order to know how many protoclusters are in the 100 lightcones at $z_p \pm \Delta z$, we select all galaxies that reside in dark matter halos that will evolve into a galaxy cluster at $z = 0$ or before. Then, we obtain the protocluster positions (right ascension, declination, and redshift) as the median of the distribution of their galaxies. Notice that the volume within a slab in our π deg^2 mocks can vary with the redshift, but the variation is small in the redshift range discussed here.

In some cases, we find clusters with less than ~ 10 galaxies in the redshift interval associated with a dark matter halo that will evolve into a massive galaxy cluster at $z = 0$. This is due to border effects (in area and/or in depth) and may produce an underestimation of the *Pure simulation* overdensities in regions occupied by protoclusters. To avoid it, we will analyze the overdensities of all protoclusters with median redshifts within the redshift slab and angular coordinates inside a radius of $1 \deg - \delta\theta/\mathrm{d_C}(\mathrm{z})$ from the center of lightcone.

Figure 4.3 shows the density contrast distributions of the field at each redshift slab (filled distributions) that represents the δ_{gal} of all pixels of the 200×200 grid inside a radius of 1 deg from the center of the lightcone. In this figure, we present both *Pure Simulation* samples in the first and the third columns, for the two surveys being emulated in this section. It is also shown the density where protocluster galaxies reside.

4.1.4 | Observational-like Overdensities

The *Observational-like* samples represent the data sets constrained by the $i \leqslant 25$ mag magnitude limit for the wide layer of the *HSC-SSP* and *LSST* surveys. The corresponding photometric redshift estimations obtained using `Le Phare` are shown in Figure 3.4.

`Le Phare` output gives the most likely redshift, z_{phot} in each case. We have done overdensity maps (see Figure 4.1) from the distribution of galaxies within $|z_{\mathrm{phot}} - z_p| \leqslant \sigma_{\mathrm{NMAD}} \times (1 + z)$, where σ_{NMAD} is 0.034 and 0.020, for the *HSC-like* and *LSST-like* samples, respectively, as derived previously.

The density contrast distributions for these *Observational-like* samples are also shown in Figure 4.3, where the filled distributions in the second and fourth columns correspond to *HSC-like* and *LSST-like* samples, respectively. Additionally, we present in each panel a strict outlier

fraction, η, using Equation 4.2:

$$\eta = \frac{N_{\Delta z}[|\delta_z|/(1+z_r) > \sigma_{\text{NMAD}}]}{N_{\Delta z}} \tag{4.2}$$

where $N_{\Delta z}$ is the number of photo-z selected galaxies within each redshift slab. This Figure also shows the galaxy surface density within the redshift slabs ($n_{\text{gal}} = N_{\Delta z}/\pi \text{ deg}^{-2}$).

4.1.5 | Protocluster Overdensities

Protoclusters are often found in observations as peaks in density contrast maps. To quantify their surface density in slabs of our survey emulations, we first extract all overdensity peaks in a density contrast map of the slab, computed in the 200×200 pixels grid covering celestial coordinates $-1.0 < \Delta RA < 1.0$ [deg] and $-1.0 < \Delta Dec < 1.0$ [deg] (therefore, each pixel corresponds to 0.6 arcmin). To identify the overdensity peaks, we have compared each pixel of these maps with the adjacent ones (3×3 pixels matrix). If the central pixel presents the highest value, then we select its position as an overdensity peak.
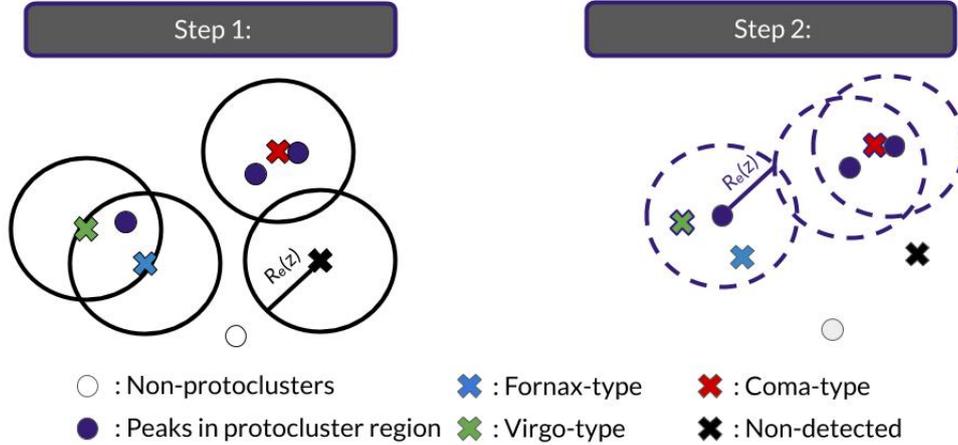
Figure 4.2: Representation of the procedure that links each overdensity peak with a specific protocluster.

After, we analyze the surroundings of all protoclusters within each redshift slab. Our first step consists in identifying all peaks in protoclusters regions, which are inside a radius equals to the bandwidth, $\delta\theta$, centred at the position of a given structure. At the end of this step, we obtain two main samples: peaks in protocluster regions and those that are not. Since we expect protoclusters to be a collection of dark matter halos, they can contain more than one overdensity peak within $\delta\theta$. Nevertheless, given our choice of kernel bandwidth (effective radius of Coma-type protoclusters at z_p), this happens for less than 15% of the structures in the entire sample.

The next step consists of linking all peaks which are associated with protocluster regions to a given structure. In this case, we search for protoclusters inside the effective radius but centered at the position of the peaks. Then, the peak is associated to the most massive structure, if there are more than one protoclusters in this area. We present this procedure in Figure 4.2.

The result of this exercise is also shown in Figure 4.3, where the black solid lines show the density contrast distributions of our *HSC-like* and *LSST-like* samples for the three types of clusters discussed here: Coma-type, Virgo-type, and Fornax-type.
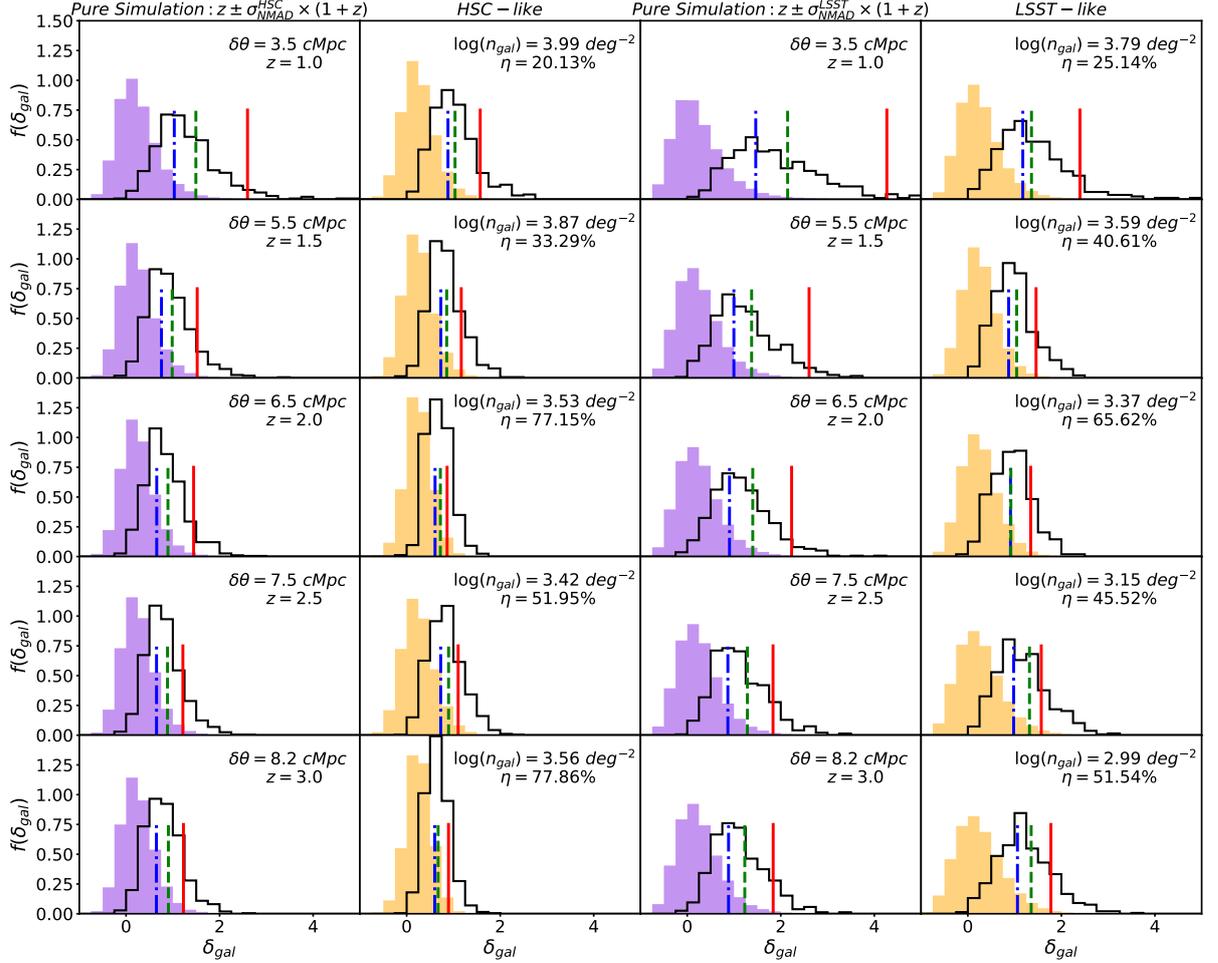
Figure 4.3: Probability density function, $f(\delta_{\mathrm{gal}})$, of density contrast, δ_{gal}, of all pixels in the maps in the π deg^2 field (filled histogram) and associated to protocluster regions (histogram with solid lines) at $z = 1.0$, 1.5, 2.0, 2.5 and 3.0 (from top to bottom) obtained from the *Pure Simulation* (first and third columns), *HSC-like* (second column) and *LSST-like* (fourth column) samples. The median overdensity of Fornax-type (blue dash-dot line), Virgo-type (green dashed line) and Coma-type (red solid line) progenitors is also presented.

4.1.6 | Protocluster Probabilities

Now we investigate what is the probability that an overdensity peak with δ_{gal} larger than a certain value is associated with a protocluster. As shown in Figure 4.3, protoclusters are found in regions with higher overdensity compared to the field, and we can use δ_{gal} to control the purity of protocluster candidates selected by overdensities.

Similar to Chiang et al. (2014) work, but by analyzing the overdensities of the peaks instead of random positions, we classify our sample into four classes: peaks in Coma, Virgo, or Fornax-type protocluster regions, and those that are not associated with any structure of this type. We designate these classes as: peaks in Coma-type (1), Virgo-type (2), Fornax-type (3) protoclusters, and in a non-protocluster region (4).

The inclusion of Coma-type progenitors in PCcones can generate biases when small volumes are analyzed (small field-of-view, or narrow redshift ranges), due to their low-density. To avoid this, we analyzed just the 80 lightcones without placed structures of the type considered at z_p. We compare the descendant $z = 0$ mass, $M_{z=0}$, distributions of the selected protoclusters within the redshift slabs of both emulated surveys with those from the $z = 0$ snapshot of the Millennium simulation. We present this comparison normalized by the volume in Figure 4.4. Since Δz is not the same for HSC and LSST emulations, the number of selected protoclusters at a certain redshift slab is not the same, as depicted in this Figure.

Additionally, the number of Coma progenitors is small; for example, we found 49 and 31 of them at $z = 1.0$ within *HSC-like LSST-like* slabs, respectively, compared with 1685 and 986 for Fornax-type protoclusters. Then, to avoid dealing with small samples, we modeled the overdensity distribution of the four different classes with a Gamma function since the fitting presents a good agreement with the measurements. We use it to generate $N \times n_i$ random overdensity values associated with each class, where N is a very large number and n_i is the volume fraction of the i-th class. We present more details of this procedure in Appendix A. Finally, we estimate the probability to find a peak with overdensity δ_{gal}, associated with the i-th class, following Equation 4.3:

$$P(\delta_{\mathrm{gal}}|i) = \frac{N_{p,i}(\geqslant \delta_{\mathrm{gal}})}{\sum N_{p,i}(\geqslant \delta_{\mathrm{gal}})},\tag{4.3}$$

where $N_{p,i}(\geqslant \delta_{\mathrm{gal}})$ is the number of peaks with overdensity higher than δ_{gal}, associated with the i-th type. Notice that the used value of $N_{p,i}$ comes from our sampling with the Gamma function. We present these probabilities as a function of δ_{gal} in Figure 4.5 for *HSC-like* and *LSST-like* mock surveys on the first and second rows, respectively, for redshifts $z = 1.0, 1.5, 2.0, 2.5$ and 3.0 (from left to right, respectively).

Figure 4.5 shows, as expected, that by adopting higher values of δ_{gal}, we can achieve a higher probability of the selected peaks to be associated with a real structure. Based on this figure, we can establish definitions to classify protocluster candidates. We restrict our sample of peaks to achieve a 70 % of confidence that they are genuine protoclusters. To obtain this overdensity lower limit, we find the value when the volume fraction of non-protoclusters drops
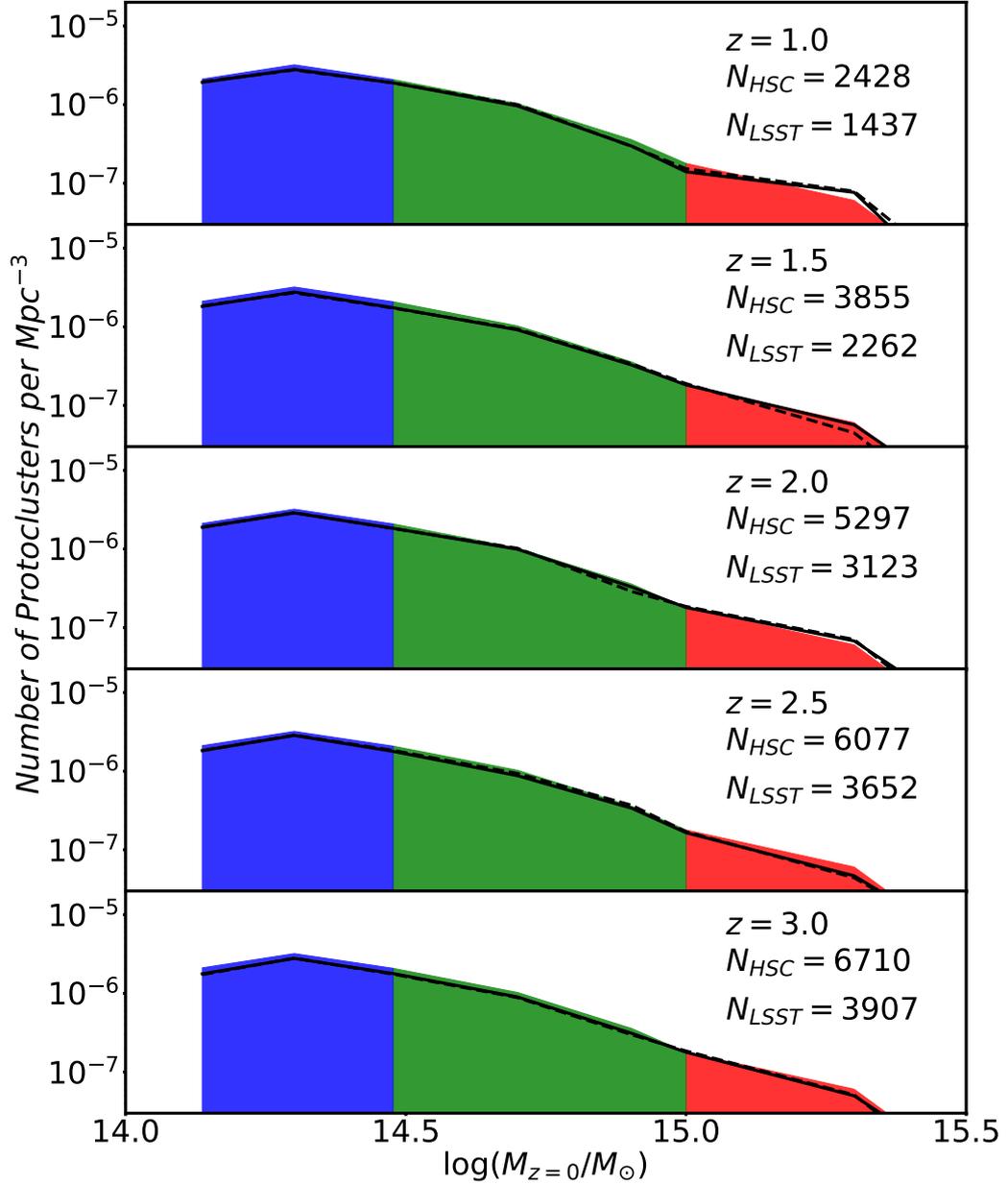
Figure 4.4: Descendant $z = 0$ cluster mass ($M_{z=0}$) distribution of the selected protoclusters per Mpc^3 at $z_p = 1.0, 1.5, 2.0, 2.5$ and 3.0, from top to bottom panels, respectively. Solid and dashed lines indicate the distribution for the *HSC-like* and *LSST-like* slabs, respectively, while filled areas represent those obtained from the $z = 0$ Millennium simulation snapshot with different colours for Fornax, Virgo and Coma-type clusters.

to 0.3, which is equivalent to the desired detection accuracy. On the other hand, by conditioning the sample with respect to δ_{gal}, we limit the total number of detected structures that are real, i.e., the completeness decreases. Additionally, as Figure 4.5 shows, the protocluster probabilities are descendant-mass dependent. Therefore, the completeness is different for Coma, Virgo, and Fornax-type protoclusters. We summarize our results of detection completeness for the adopted criterium in Table 4.1 for *HSC-like* and *LSST-like* samples for different redshift slabs.
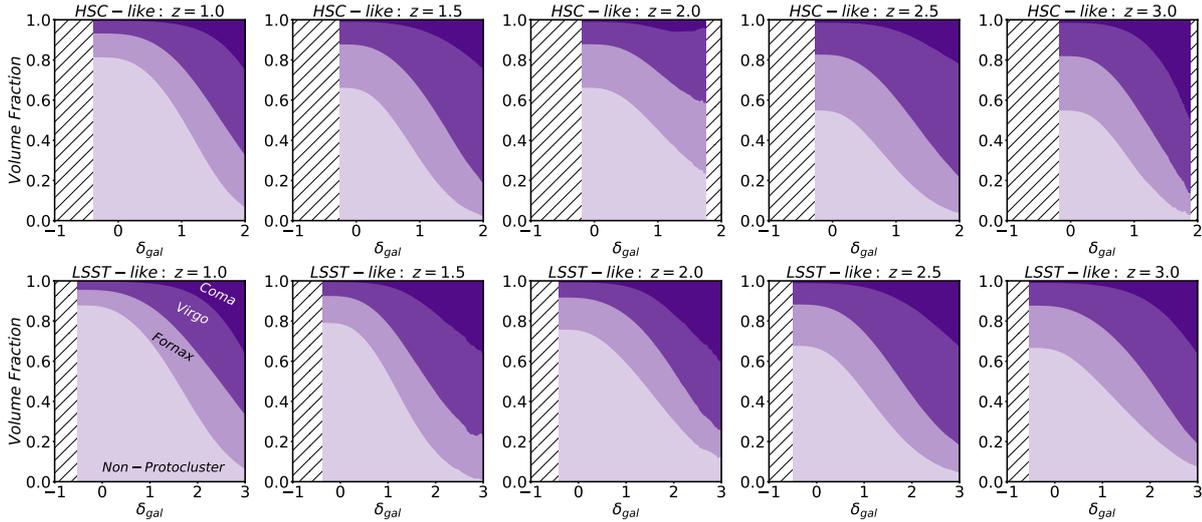
Figure 4.5: Probability of a peak with overdensity δ_{gal} be a non-protocluster (light purple), Fornax-type, Virgo-type, or Coma-type (dark purple) protocluster. These probabilities are based on overdensity estimations using a Gaussian KDE with bandwidth $\delta\theta$ (see Table 4.1), at $z = 1.0$, 1.5, 2.0, 2.5 and 3.0, from left to the right. The first and second rows present results for the *HSC-like* and *LSST-like* samples. Hatched regions represent overdensity ranges without peaks.

Table 4.1: Predicted galaxy overdensity, δ_{gal}, required to have 70% of probability of being a real protocluster. We also show the expected completeness associated to this criteria for the full sample (C_{all}), Coma (C_{C}), Virgo (C_{V}), and Fornax type (C_{F}) protoclusters, as well as the kernel bandwidth $\delta\theta$.

		HSC-like					*LSST-like*			
z	$\delta\theta$ [arcmin]	δ_{gal}	C_{all} [%]	C_{C} [%]	C_{V} [%]	C_{F} [%]	δ_{gal}	C_{all} [%]	C_{C} [%]	C_{V} [%]	C_{F} [%]
1.0	3.53	1.39	11.45	61.22	17.72	7.42	2.08	8.28	54.84	11.67	5.38
1.5	4.21	1.01	13.00	48.65	19.87	9.20	1.54	5.17	41.46	7.26	3.37
2.0	4.20	1.48	0.21	0.92	0.34	0.13	2.02	0.93	8.57	1.32	0.54
2.5	4.31	0.89	18.18	58.33	29.16	12.90	1.49	13.94	63.64	25.37	8.11
3.0	4.33	0.93	6.23	29.84	10.56	3.90	1.72	10.06	51.95	18.06	5.77

4.1.7 | Cluster detectability

In this section, we focus on the 20 protoclusters placed at $z = 1.0$, 1.5, 2.0, 2.5, and 3.0 (Section 2.8). We summarize the overdensity measurements of these structures in Table 4.2, for the two surveys. Protoclusters that do not appear in the table are those without an overdensity peak within a projected distance $R_e(z)$ of their centers. For the HSC-SSP simulation we have

detected, at these redshifts, 16, 14, 8, 14, and 11 protoclusters, respectively. For the LSST forecast, the corresponding numbers are 18, 15, 12, 14 and 13, respectively.

Table 4.2: Detected overdensities of the 20 protoclusters for the *HSC-like* and *LSST-like* samples at $z = 1.0, 1.5, 2.0, 2.5$ and 3.0.

Descendant Type	$\log (M_{z=0}/M_\odot)$	δ_{gal} (*HSC-like*)					δ_{gal} (*LSST-like*)				
		$z = 1.0$	$z = 1.5$	$z = 2.0$	$z = 2.5$	$z = 3.0$	$z = 1.0$	$z = 1.5$	$z = 2.0$	$z = 2.5$	$z = 3.0$
Fornax	14.14	0.51	0.43	—	0.58	—	1.15	—	—	0.75	—
	14.17	1.37	—	—	1.45	0.53	1.77	1.17	1.45	—	2.32
	14.20	1.20	—	—	—	1.01	1.54	1.62	—	—	1.18
	14.23	—	0.67	—	0.72	—	0.76	0.77	0.97	1.22	1.46
	14.27	0.88	0.86	—	1.02	—	1.46	0.86	—	—	1.43
	14.31	1.22	0.93	—	—	0.61	1.77	0.75	0.64	0.85	1.81
	14.36	—	0.53	0.72	1.01	—	0.98	0.42	—	0.95	—
	14.42	0.50	—	0.34	—	0.87	0.34	—	0.93	—	—
Virgo	14.49	—	—	—	1.27	—	—	1.10	—	—	—
	14.58	1.33	0.98	0.66	0.56	—	1.43	—	1.09	1.27	1.12
	14.73	1.09	0.83	0.75	0.76	—	1.69	1.20	—	1.20	1.81
	14.73	1.31	1.07	0.74	1.31	0.66	1.55	1.44	1.17	1.62	1.64
	14.82	1.42	1.17	0.63	1.03	0.84	2.16	1.10	—	1.79	1.76
	14.94	1.34	0.91	—	—	—	2.44	0.81	0.47	—	—
Coma	15.03	1.31	0.91	—	0.90	0.46	2.55	—	0.67	1.48	1.50
	15.05	2.22	1.90	—	—	—	3.02	2.21	1.66	2.18	—
	15.07	1.10	—	—	—	1.00	1.87	—	1.34	1.97	1.82
	15.16	1.09	—	—	1.30	1.03	2.12	1.15	—	2.49	2.04
	15.26	—	1.55	0.87	1.51	0.85	—	1.67	1.65	1.74	2.29
	15.36	2.61	1.59	1.09	1.12	0.74	4.40	1.64	1.60	2.22	—

Considering Coma-type protoclusters, we recovered at least 4 out of 6 systems; an exception is for $z = 2.0$ for the *HSC-like* case, where only 2 out of 6 were detected. This is due mainly to photo-z uncertainties, since at $z = 2.0$, many protocluster members (independently of the cluster type) have a measured photo-z outside the redshift slab for both simulated surveys. This can be seen in Figure 4.6, where we show the photometric redshift distribution of member galaxies of a Coma-like non-detected protocluster ($\log (M_{z=0}/M_\odot) = 15.16$) for both survey emulations. Therefore, photometric redshift uncertainties can dilute real overdensities.
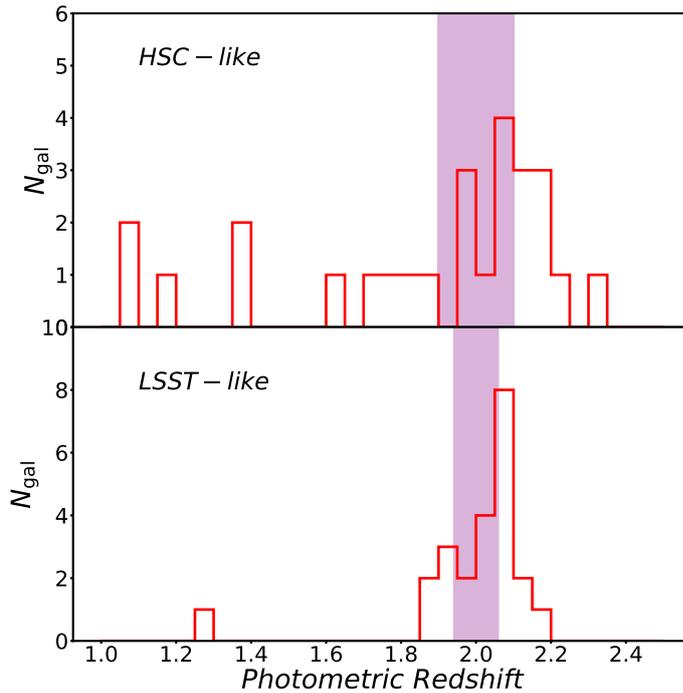
Figure 4.6: Photometric redshifts distribution (*HSC-like* and *LSST-like* photo-zs in the top and bottom panels, respectively) of the Coma-type protocluster with $\log\left(M_{z=0}/M_{\odot}\right) = 15.16$ which was not detected. The shaded region represents the redshift slab.

It is not the case for the non-detected Coma-type at $z = 1.0$, however. This particular structure is marked with a red circle in Figure 4.1, and it is, in fact, in a dense environment, since it concentrates ~ 77 photo-z selected members in the main dark matter halo within a radius of 1.98 arcmin and is ~ 3.3 times denser than the field. Nevertheless, the detected overdensity peak suffered a shift due to the projection of galaxies in secondary halos, some of them members of a massive group ($\log\left(M_{z=0}/M_{\odot}\right) = 13.70$). In conclusion, both photometric redshift uncertainties and projection effects may preclude even the detection of rich structures.

4.2 | Analysis of the CFHTLS

We have been focusing on protocluster detection at specific redshifts. In this section, we emulate continuous cluster detection at $z \leqslant 1.5$, using the *CFHT-like* sample that simulate the Deep Canada France Hawaii Telescope Legacy Survey (CFHTLS; http://www.cfht.hawaii.edu/Science/CFHTLS/). To do this, we compute the overdensity maps at 29 different overlapping redshift slabs centered at $z = 0.1$ to 1.5 and spaced by redshift intervals of 0.05. We have selected all galaxies with photo-z within $z_i \pm \Delta z = z_i \pm \sigma_{\mathrm{NMAD}}^{CFHT} \times (1 + z)$, where $\sigma_{\mathrm{NMAD}}^{CFHT} = 0.027$ and z_i is the center of each slab. We use in this case, for simplicity, a Gaussian kernel with bandwidth $\delta\theta = 1.0$ Mpc. Therefore, the density field here is described by a cube composed by 29 slabs. We apply a procedure similar to that described earlier to extract overdensity peaks, but now

using a $3 \times 3 \times 3$ pixels analyser cube (corresponding to 0.6×0.6 in arcmin and 0.05 in redshift) that runs over all the 29 maps data cube. Hence, if the central pixel of the analyzer cube is a maximum, then we register its position.

We repeat this method for 6 PCcones without putting *a priori* any structures within the redshift interval of interest because it could add bias to the cluster distribution.

Finally, we define as cluster candidates all overdensity peaks with $\delta_{\mathrm{gal}} \geqslant 2.0$. This value is the same density contrast cut used by Durret et al. (2011). It corresponds to median(δ_{gal})+std(δ_{gal}), where median(δ_{gal}) and std(δ_{gal}) are the median and the standard deviation of δ_{gal}, respectively. However, since we are using a fixed bandwidth in Mpc, maps at high redshift present a large number of overdense peaks, because $\delta\theta$ decreases in angular coordinates, the number of galaxies decreases, and these two factors tend to amplify the density contrast. Moreover, these peaks can enclose just a few galaxies within a radius of 1 Mpc. To avoid these statistical fluctuations we establish a minimum number of galaxies inside a radius of 1 Mpc, $N_{min} = 17$, which represents the median number of galaxies of overdensity peaks with $\delta_{\mathrm{gal}} \geqslant 2.0$. The cumulative redshift distribution for the mock cluster candidates is shown in Figure 4.7 (black solid line). Figure 4.7 also presents the cluster distribution obtained by Wen and Han (2011) (green dashed line) for the Deep CFHTLS. The Deep CFHTLS is a 3.2 deg^2 field with 5σ magnitude limits presented in Table 3.1. Ilbert et al. (2006) and Coupon et al. (2009) obtained redshift uncertainties $\sigma_{\Delta z}/(1 + z) = 0.029$ for galaxies with $i < 24.0$ mag for this field. Following Wen and Han (2011), we adopted here the same magnitude limit. We have performed a *Kolmogorov-Smirnov* test to verify whether the Wen and Han (2011) sample and our simulated cluster distribution are consistent with each other. We obtained a $p - value = 18.4\%$, indicating that both distributions are indeed similar.

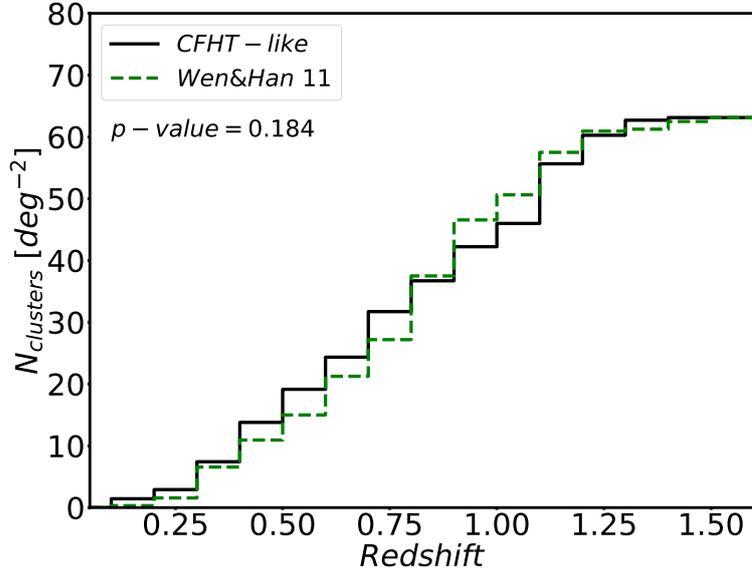Figure 4.7: Cumulative cluster per deg^2, N_{clusters} per redshift bin, from the cluster detection of Wen and Han (2011) (green dashed line) for the Deep CFHTLS. Our results from the deep *CFHT-like* sample are shown as a black continuous line.

5 Discussion

The detection of structures at high redshifts, such as galaxy clusters or protoclusters, is critically affected by observational constraints. We have explored in this work the impact of magnitude constraints and photometric redshift estimations on the analysis of the galaxy density field derived from photometric surveys by using a new implementation of mock catalogs that we call PCcones. The particularity of PCcones is that a selected structure is placed at the center of each mock. This increases the number of massive protoclusters at specific redshifts, which is low due to their rarity. These mocks facilitate analyzes about the physical properties of members in different protocluster types and how observational constraints affect these results. This will be explored in future works.

Additionally, we have adopted a different procedure to obtain apparent magnitudes for the Millennium Lightcone that uses the post-processing technique of Shamshiri et al. (2015). This is done by attributing a SED to all mock galaxies from their star formation histories, computed with the `L-GALAXIES` semi-analytic model of Henriques et al. (2015).

In Chapter 3 we have presented results on photometric redshift estimations for PCcones and the Millennium Lightcone, using the `Le Phare` SED fitting algorithm. We notice that our approach improves these estimations due to the method to obtain observer-frame magnitudes. For instance, the Millennium Lightcone presents a clear bias at $z \lesssim 1.0$, even considering that both mocks were computed using the same SAM. Recently, Laigle et al. (2019) presented predictions for photometric redshifts and physical properties of galaxies for the LSST survey by using mock catalogs based on hydrodynamical simulations. Laigle et al. (2019) also attributed a SED to each simulated galaxy to obtain apparent magnitudes. Their results for photo-zs are similar to ours. For example, for galaxies in the i band bin $22 < i \leqslant 23$ mag, we obtain $\sigma_{\mathrm{NMAD}} = 0.017$ and an outlier fraction $f_{\mathrm{outliers}} = 0.1\%$, which are exactly the same values obtained by Laigle et al. (2019). However, for deeper i band magnitude bins, our photo-z predictions are more accurate. These differences can be due be due to the differences in the simulations and in the IGM absorption model (they applied the IGM absorption to the galaxy SEDs according with the gas distribution in the simulated IGM). We have also shown, in Section 3.3, that our Deep *CFHT-like* mock presents a difference from the observational results of Ilbert et al. (2006) of only about 2%. Our results in that section indicate that PCcones reproduce reliably observed photo-z estimates without the implementation of additional codes for estimation of mock magnitudes such as `PhotReal` (Ascaso et al., 2015). We can accurately emulate surveys/observations by adding the respective noise to PCcones magnitudes. This is useful to test different approaches to photo-z estimation and structure detection.

It is a common practice, in photometric surveys to look for structures in redshift slabs and, in this case, the slab width plays a critical role. Chiang et al. (2013) has examined the impact of redshift uncertainties on protocluster overdensities, noticing that with a slab width $\Delta z = 0.15$,

many random regions could be spurious candidates.

We notice that, as expected, the distribution of the density contrast, δ_{gal}, of the field (filled histogram in Figure 4.3) is different between the *Pure Simulation* and *Observational-like* samples, mainly when the number of galaxies in the field is low or the outlier factor η is high.

In Figure 4.5, we present the probability of an overdensity peak in our maps be a real structure. This figure help us to define a criterion to classify protocluster candidates based on δ_{gal}. We have analyzed protocluster detectability when we expect 70% of purity. To find the lower-limit of overdensity to restrict the sample of peaks, we obtain the value when the non-protocluster distribution drops to 0.3. Similarly, other criteria can be adopted, for example, to reach 60% or 80% of selection accuracy.

We also present in Figure 4.5 the volume fraction of different cluster type progenitors as a function of the overdensity δ_{gal}. As expected, Coma-type protoclusters are associated with high overdensities. However, we found a non-negligible fraction of Virgo and Fornax-type protoclusters with very high overdensities. This is due to projection effects, where different independent structures in the same area amplify the overdensities of low mass protoclusters. This effect is more noticeable in the *HSC-like* sample because the redshift slab is larger. We found that the mean of the number of protoclusters within the effective radius matched with Virgo-types is $\gtrsim 2.3$ at the redshifts where this trend is stronger.

From Table 4.1, we show that the completeness at 70% confidence level for *HSC-like* is higher than that obtained for the *LSST-like* sample in almost all redshifts. This does not imply, however, that structure detection for the wide HSC-SSP is better than the expected for LSST, whose photo-z estimations are more accurate. If we reanalyze the *LSST-like* sample with the same slab width used for the *HSC-like* sample, we find that C_{all} for *LSST-like* is in general higher or equal to that obtained for *HSC-like*. This occurs because for higher redshift slabs there are more structures in the same area (see Figure 4.4), and since the number of overdensity peaks in both fields are similar, the chance to match protoclusters with overdensity peaks increases for the *HSC-like* sample.

From Figure 4.5 and mainly in Table 4.1, we evidence that structure detectability presents a strong dependence with redshift, mostly due to the variation in the accuracy of photo-zs. This suggests that detection criteria should be established as a function of redshift. Notice that the completeness of the protocluster sample is lower when η is high ($\gtrsim 60\%$). This occurs at $z = 2.0$ and 3.0 for the *HSC-like* sample, where the achieved completeness is 0.21% and 6.23%, respectively. It also happens at $z = 2.0$ in *LSST-like* emulation, where the expected completeness is 0.93%.

The magnitude limit plays an important role in overdensity estimations due to the *Malmquist-bias*, which affects galaxy completeness. The construction of density contrast maps with low galaxy completeness leads to shot-noise, where a small set of galaxies generates overdense regions (e.g. Vicentin et al., submitted). However, this does not happen in our case given the imposed magnitude limit ($i = 25.0$ mag)

Comparing our results with other similar works, Chiang et al. (2014) obtained a sample of protocluster candidates by using photometric redshifts from the COSMOS/UltraVista catalog (Muzzin et al., 2013), whose photo-z uncertainty is $\sigma_z = 0.025(1 + z)$. They adopted as detection

criterion the mean overdensity of Coma-type protoclusters, based on lightcones. In overall, they achieved a purity of $\sim 70\%$, and completeness of 9%, 7%, 17%, and 50% for all, Fornax, Virgo, Coma-type structures, respectively, at $1.6 < z < 3.1$. Considering redshift slabs at $1.5 \leqslant z \leqslant 3.0$, we obtain, on average, completeness of $\sim 7\%$, 4%, 13%, and 40% for the same classes, respectively, in the *LSST-like* sample. Ando et al. (2020) constructed another potential protocluster candidates catalog by searching for pairs of massive galaxies in the COSMOS2015 data (Laigle et al., 2019) at $z \sim 2.0$, which presents one of the most accurate photo-zs in the literature ($\sigma_z \sim 0.007(1 + z)$). Given their selection criteria, they estimated that 54% of the observed galaxy pairs are real and that 63% of them reside in protocluster regions ($M_{z=0} > 10^{14}~M_\odot$). From simulations, they estimate completeness of 23%, 16%, 52%, and 100% for the whole, Fornax, Virgo, and Coma-type protocluster samples, respectively. At the same confidence level, for the *HSC-like* sample, we achieve that the corresponding completenesses are $\sim 31\%$, 26%, 42%, and 64%.

Despite all the factors that affect protocluster detection in photometric surveys and the simplicity of our method, we obtain acceptable results for our HSC-SSP and LSST emulations. For example, if we search for protocluster candidates within redshift slabs of width $z_p \pm \sigma_{\mathrm{NMAD}} \times (1 + z_p)$, and we set the kernel bandwidth as listed in Table 4.1, we can achieve completeness ranging from $\sim 12\% - 18\%$ for a 70% confidence level at $z_p = 1.0$, 1.5 and 2.5 for the *HSC-like* sample. In the case of the *LSST-like* sample, our best results are at $z = 2.5$ and 3.0; here, we can recover $13\% - 18\%$ of all protoclusters with 70% of purity in a narrower redshift slab. To reach these results, we have to apply a cut in the overdensities (see δ_{gal} in Table 4.1) of the peaks in the fields at the considered redshifts. With our definition of candidates, we can detect, on average, $\sim 39\%$ and 43% of the Coma-type progenitors for the *HSC-like* and *LSST-like* samples, respectively at $1.0 \leqslant z \leqslant 3.0$. Given the expected number of Coma-type protoclusters within $1.0 < z < 3.0$ in each survey sky coverage (1,400 deg^2 and 20,000 deg^2, for HSC-SSP and LSST, respectively), we would be able to detect $\sim$3033 and $\sim$43,329 progenitors of the most massive clusters in the full HSC-SSP and LSST areas, by using our simple technique.

6 Conclusions and Future Perspectives

We divide this final chapter into two sections: conclusions and future perspectives. In the former, we present a summary and conclusion of this work. In the second part, we present some ideas that we plan to develop during the PhD.

6.1 | Conclusions

In this MSc Dissertation we have introduced a new statistical tool to analyze the level of significance of overdensity estimations associated to the search of protoclusters of galaxies. Our main results are as follows:

1. We have developed a procedure that allows to build a mock lightcone that contains some desired structure at the center of the line-of-sight and at a certain redshift of interest. These PCcones have been constructed by redefining the mock zero-point (i.e., $z = 0$). The zero-point is important to place correctly the center of masses of a protocluster at a given redshift z_p (see Section 2.2).

2. We have estimated mock apparent magnitudes at a *post-processing* stage, using the SFH arrays generated by the `L-GALAXIES` semi-analytic model of Henriques et al. (2015). Galaxy SEDs have allowed us to obtain reliable observer-frame magnitudes, without using magnitude interpolations or K-corrections (see Section 2.5).

3. We have estimated photometric redshifts in mock surveys (see Chapter 3), using `Le-Phare`. To do this, we have created *Observational-like* mocks, emulating the Deep CFHTLS (*CFHT-like*), the wide layer of HSC-SSP (*HSC-like*), and the 10 years forecast of LSST (*LSST-like*), by applying observational constraints on the photometric bands of each of these surveys, accordingly to their 5σ magnitude limits.

4. The comparison of our mock photometric redshifts with their observational counterpart for the *CFHT-like* and *HSC-like* samples gives satisfactory results (see Figure 3.4). The comparison of our photo-z predictions are consistent with other LSST photo-z forecasts. Additionally, we have estimated photo-zs using the Millennium Lightcone from Henriques et al. (2015) for the *CFHT-like* sample, finding that the Millenium photometric redshifts present a clear bias at $z \lesssim 1.0$ (see Figure 3.2), contrarily to what we have obtained. We attribute this difference to the procedure we adopted to compute the apparent magnitudes.

5. We have constructed galaxy density contrast maps for the PCcones at five different redshifts ($z = 1.0$, 1.5, 2.0, 2.5 and 3.0), using a two-dimensional Gaussian kernel (Figure 4.1 shows an example for a protocluster at $z = 1.0$). We have focused on overdensity estimations for two types of samples: the *Observational-like* sample, composed by the *HSC-like* and *LSST-like* mock surveys, and, for comparison, the *Pure Simulation* sample, without any observational/magnitude constraints. For both samples, we have obtained density maps in slabs using our estimated photometric redshifts.

6. For each density contrast map, we extracted all overdensity peaks within a redshift slab and investigated their association with protoclusters. This allowed us to estimate the probability of protocluster detection at a given density contrast δ_{gal}, as well as the completeness of this sample at a 70% confidence level. Our main results are summarized in Table 4.1. For the HSC-SSP emulation, we expect to recover $\sim 12\% - 19\%$ of all structures at $z_p = 1.0$, 1.5, and 2.5. For the 10-year forecast of the LSST, within a narrower redshift slab, these numbers are between $8\% - 14\%$ at $z_p = 1.0$, 2.5 and 3.0.

7. In some cases, the combination of observational constraints and photo-z uncertainties affect the detection of structures critically. This happens at $z_p = 2.0$ for both emulated samples. We found that the completeness of all protoclusters is $\sim 1\%$, given our selection criterion, for this redshift.

8. Table 4.1 shows the that structure detectability (value of δ_{gal} required to achieve 70% of purity, and completeness associated to this criterion) changes drastically with redshift, mostly due to the decreasing accuracy of photo-zs. This suggests that selection criteria should be established as a function of redshift.

9. Finally, we have emulated cluster detection in a *CFHT-like* sample, obtaining cluster distributions similar to actual observations.

6.2 | Future Perspectives

Our main topics of research are mock construction and protocluster detection. During this project, we have dealt with both satisfactorily. For example, with our lightcones, we can reproduce reliable observed photo-z measurements. Also, with our implemented method to detect structures, we achieved similar performance than previous works. However, we can perform many improvements in these issues.

6.2.1 | Upgrade PCcones

Our PCcones present consistency with observational data in optical broad-band photometry, which is enough for this work. But, this tendency is not maintained for near and mid-infrared. We have compared colors and magnitudes with the Laigle et al. (2019) lightcone and COSMOS2015 data. Our results show that the three distributions differ with each other considerably for redder filters (from VISTA-J to IRAC4.5μm bands). To deal with this, we can perform several tests with different SEDs templates of stellar population synthesis models.

If we want to extend the usage of our mocks for longer wavelengths, we have to link dust emission models with simulated galaxies. A reasonable approximation can be by attributing dust emission SED according to their physical properties. Examples of these templates include Draine and Li (2007), Casey (2012), and Dale et al. (2014).

Another upgrade to our lightcones is the addition of emission lines to galaxy SEDs. Izquierdo-Villalba et al. (2019) included flux contribution of 9 emission lines, following Orsi et al. (2014) model to J-PLUS mock. The modeled lines were Lyα (1216 Å), Hβ (4861 Å), Hα (6563 Å), [O II] (3727 Å, 3729 Å), [O III] (4959 Å, 5007 Å), [Ne III] (3870 Å), [O I] (6300 Å), [N II] (6548 Å, 6583 Å), and [S II] (6717 Å, 6731 Å). With this, we will be able to emulate surveys/observations with narrow-band photometry.

A current main question is what we are tracing if we use different selection criteria of high redshift galaxies. It is a fact that the extraction of LAEs and LBGs pinpoint distinct structures (e.g. Bădescu et al., 2017; Shi et al., 2019). Even more, observing in longer wavelengths, for example, Rotermund et al. (2020) found no more than 4 LBGs surrounding the Coma-type protocluster core at $z = 4.3$ reported by Miller et al. (2018). The inclusion of all these improvements in PCcones will allow us to analyze this problem in a theoretical context. As well, they will extend the usefulness of lightcones in interpretations about detected protoclusters besides broad-band optical photometry.

6.2.2 | Protocluster Detection

We have focused on protocluster detection by using photometric redshifts. Although our implemented method is simple, we achieve results similar to other works (e.g. Chiang et al., 2014; Ando et al., 2020). However, we can upgrade our technique by developing a more robust approach.

We can adopt a probabilistic approximation to select galaxies at a certain redshift interval, instead of by using their most likely values. To do this requires the full redshift probability density distribution (PDF). In practice, we would compute density contrast maps with all galaxies in the field, but with an associated weight equals to the probability of the source to be at the redshift

slab. With this, we expect to reduce the impact of photo-z uncertainties over protocluster detection. This approximation is not new, since many algorithms use the full PDF information to detect structures, such as the Bayesian Cluster Finder (Ascaso et al., 2012), `PZWav` (Gonzalez, 2014), `WaZP` (Benoist, 2014; Dietrich et al., 2014), `AMICO` (Bellagamba et al., 2018), among others. We can also add other priors in the algorithm, such as the mean redshift distribution, colors and luminosity functions.

Moreover, we constructed density contrast maps with a Gaussian Kernel Estimation. This type of kernel complicates the analysis of border effects since all data points contribute to the density contrast at any angular position. However, many kernel functions could be explored, such as Epanechnikov, quartic, tri-weight, tri-cube, among others, which presents a delimited area given by the kernel bandwidth. In this work, we define this parameter as the effective radius of Coma-type protoclusters at redshift z_p. As well, we should test different choices of this quantity.

Finally, we plan to apply our methods to detect new high-z structure candidates in observational data. At present, over $300 \deg^2$ were observed by HSC-SSP. This makes it a perfect target for new protocluster discoveries. After producing a catalog of protocluster candidates, we can perform spectroscopic follow-ups of the most interesting candidates to confirm them and investigate their properties. Two of the main field under observation by HSC-SSP are between declination ± 5 deg, allowing us to request observing time for telescopes in both hemispheres.

Bibliography

Adams, S. M., Martini, P., Croxall, K. V., Overzier, R. A. and Silverman, J. D., "Discovery of an overdensity of Lyman alpha emitters around a z $\sim$ 4 QSO with the Large Binocular Telescope", MNRAS, vol. 448, no. 2, pp. 1335–1344, 2015. 1407.2609, URL http://dx.doi.org/10.1093/mnras/stv065.

Aihara, H. et al., "The Hyper Suprime-Cam SSP Survey: Overview and survey design", PASJ, vol. 70, S4, 2018. 1704.05858, URL http://dx.doi.org/10.1093/pasj/psx066.

Almosallam, I. A., Lindsay, S. N., Jarvis, M. J. and Roberts, S. J., "A sparse Gaussian process framework for photometric redshift estimation", MNRAS, vol. 455, no. 3, pp. 2387–2401, 2016. 1505.05489, URL http://dx.doi.org/10.1093/mnras/stv2425.

Ando, M., Shimasaku, K. and Momose, R., "A systematic search for galaxy proto-cluster cores at z $\sim$ 2", MNRAS, vol. 496, no. 3, pp. 3169–3181, 2020. 2002.10511, URL http://dx.doi.org/10.1093/mnras/staa1757.

Angulo, R. E. and White, S. D. M., "One simulation to fit them all - changing the background parameters of a cosmological N-body simulation", MNRAS, vol. 405, no. 1, pp. 143–154, 2010. 0912.4277, URL http://dx.doi.org/10.1111/j.1365-2966.2010.16459.x.

Araya-Araya, P. et al., "Protocluster detection in simulations of HSC-SSP and the 10-year LSST forecast, using PCcones", , submitted.

Arnouts, S. et al., "ESO imaging survey. Deep public survey: Multi-color optical data for the Chandra Deep Field South", A&A, vol. 379, pp. 740–754, 2001. astro-ph/0103071, URL http://dx.doi.org/10.1051/0004-6361:20011341.

Arnouts, S. et al., "Measuring the redshift evolution of clustering: the Hubble Deep Field South", MNRAS, vol. 329, no. 2, pp. 355–366, 2002. astro-ph/0109453, URL http://dx.doi.org/10.1046/j.1365-8711.2002.04988.x.

Ascaso, B., Mei, S. and Benítez, N., "Apples to apples A^2 - I. Realistic galaxy simulated catalogues and photometric redshift predictions for next-generation surveys", MNRAS, vol. 453, no. 3, pp. 2515–2532, 2015. 1503.01113, URL http://dx.doi.org/10.1093/mnras/stv1597.

Ascaso, B., Wittman, D. and Benítez, N., "Bayesian cluster finder: clusters in the CFHTLS Archive Research Survey", MNRAS, vol. 420, no. 2, pp. 1167–1182, 2012. 1111.0013, URL http://dx.doi.org/10.1111/j.1365-2966.2011.20107.x.

Ascaso, B. et al., "An accurate cluster selection function for the J-PAS narrow-band wide-field survey", MNRAS, vol. 456, no. 4, pp. 4291–4304, 2016. 1601.00656, URL http://dx.doi.org/10.1093/mnras/stv2988.

Balogh, M. L. et al., "Gemini Observations of Galaxies in Rich Early Environments (GOGREEN) I: survey description", MNRAS, vol. 470, no. 4, pp. 4168–4185, 2017. 1705.01606, URL http://dx.doi.org/10.1093/mnras/stx1370.

Baugh, C. M., "A primer on hierarchical galaxy formation: the semi-analytical approach", *Reports on Progress in Physics*, vol. 69, no. 12, pp. 3101–3156, 2006. astro-ph/0610031, URL http://dx.doi.org/10.1088/0034-4885/69/12/R02.

Bellagamba, F., Roncarelli, M., Maturi, M. and Moscardini, L., "AMICO: optimized detection of galaxy clusters in photometric surveys", MNRAS, vol. 473, no. 4, pp. 5221–5236, 2018. 1705.03029, URL http://dx.doi.org/10.1093/mnras/stx2701.

Benítez, N., "Bayesian Photometric Redshift Estimation", ApJ, vol. 536, no. 2, pp. 571–583, 2000. astro-ph/9811189, URL http://dx.doi.org/10.1086/308947.

Benoist, C., "wazp cluster finder", in: *Building the Euclid Cluster Survey - Scientific Program*, p. 8 (2014).

Bertone, G. and Tait, T. M. P., "A new era in the search for dark matter", Nature, vol. 562, no. 7725, pp. 51–56, 2018. 1810.01668, URL http://dx.doi.org/10.1038/s41586-018-0542-z.

Bianconi, M. et al., "LoCuSS: pre-processing in galaxy groups falling into massive galaxy clusters at z = 0.2", MNRAS, vol. 473, no. 1, pp. L79–L83, 2018. 1710.04230, URL http://dx.doi.org/10.1093/mnrasl/slx167.

Bilicki, M. et al., "Photometric redshifts for the Kilo-Degree Survey. Machine-learning analysis with artificial neural networks", A&A, vol. 616, A69, 2018. 1709.04205, URL http://dx.doi.org/10.1051/0004-6361/201731942.

Bisigello, L. et al., "The Impact of JWST Broadband Filter Choice on Photometric Redshift Estimation", ApJS, vol. 227, no. 2, 19, 2016. 1605.06334, URL http://dx.doi.org/10.3847/0067-0049/227/2/19.

Blaizot, J. et al., "MoMaF: the Mock Map Facility", MNRAS, vol. 360, no. 1, pp. 159–175, 2005. astro-ph/0309305, URL http://dx.doi.org/10.1111/j.1365-2966.2005.09019.x.

Blumenthal, G. R., Faber, S. M., Primack, J. R. and Rees, M. J., "Formation of galaxies and large-scale structure with cold dark matter.", Nature, vol. 311, pp. 517–525, 1984. URL http://dx.doi.org/10.1038/311517a0.

Böhringer, H. et al., "The ROSAT-ESO Flux Limited X-ray (REFLEX) Galaxy cluster survey. V. The cluster catalogue", A&A, vol. 425, pp. 367–383, 2004. astro-ph/0405546, URL http://dx.doi.org/10.1051/0004-6361:20034484.

Boris, N. V. et al., "Searching High-Redshift Large-Scale Structures: Photometry of Four Fields around Quasar Pairs at z ~1", ApJ, vol. 666, no. 2, pp. 747–756, 2007. astro-ph/0703562, URL http://dx.doi.org/10.1086/519992.

Brammer, G. B., van Dokkum, P. G. and Coppi, P., "EAZY: A Fast, Public Photometric Redshift Code", ApJ, vol. 686, no. 2, pp. 1503–1513, 2008. 0807.1533, URL http://dx.doi.org/10.1086/591786.

Bruzual, G. and Charlot, S., "Stellar population synthesis at the resolution of 2003", MNRAS, vol. 344, no. 4, pp. 1000–1028, 2003. astro-ph/0309134, URL http://dx.doi.org/10.1046/j.1365-8711.2003.06897.x.

Bădescu, T. et al., "Discovery of a Protocluster Associated with a Lyα Blob Pair at z = 2.3", ApJ, vol. 845, no. 2, 172, 2017. 1708.00447, URL http://dx.doi.org/10.3847/1538-4357/aa8220.

Butcher, H. and Oemler, J., A., "The evolution of galaxies in clusters. I. ISIT photometry of Cl 0024+1654 and 3C 295.", ApJ, vol. 219, pp. 18–30, 1978. URL http://dx.doi.org/10.1086/155751.

Calzetti, D. et al., "The Dust Content and Opacity of Actively Star-forming Galaxies", ApJ, vol. 533, no. 2, pp. 682–695, 2000. astro-ph/9911459, URL http://dx.doi.org/10.1086/308692.

Capak, P. et al., "A Deep Wide-Field, Optical, and Near-Infrared Catalog of a Large Area around the Hubble Deep Field North", AJ, vol. 127, no. 1, pp. 180–198, 2004. astro-ph/0312635, URL http://dx.doi.org/10.1086/380611.

Casey, C. M., "Far-infrared spectral energy distribution fitting for galaxies near and far", MNRAS, vol. 425, no. 4, pp. 3094–3103, 2012. 1206.1595, URL http://dx.doi.org/10.1111/j.1365-2966.2012.21455.x.

Chabrier, G., "Galactic Stellar and Substellar Initial Mass Function", PASP, vol. 115, no. 809, pp. 763–795, 2003. astro-ph/0304382, URL http://dx.doi.org/10.1086/376392.

Champagne, J. B. et al., "No Evidence for Millimeter Continuum Source Overdensities in the Environments of z ≳ 6 Quasars", ApJ, vol. 867, no. 2, 153, 2018. 1809.10178, URL http://dx.doi.org/10.3847/1538-4357/aae396.

Charlot, S. and Fall, S. M., "A Simple Model for the Absorption of Starlight by Dust in Galaxies", ApJ, vol. 539, no. 2, pp. 718–731, 2000. astro-ph/0003128, URL http://dx.doi.org/10.1086/309250.

Chiang, Y.-K., Overzier, R. and Gebhardt, K., "Ancient Light from Young Cosmic Cities: Physical and Observational Signatures of Galaxy Proto-clusters", ApJ, vol. 779, no. 2, 127, 2013. 1310.2938, URL http://dx.doi.org/10.1088/0004-637X/779/2/127.

Chiang, Y.-K., Overzier, R. and Gebhardt, K., "Discovery of a Large Number of Candidate Protoclusters Traced by ~15 Mpc-scale Galaxy Overdensities in COSMOS", ApJ, vol. 782, no. 1, L3, 2014. 1312.4747, URL http://dx.doi.org/10.1088/2041-8205/782/1/L3.

Chiang, Y.-K. et al., "Surveying Galaxy Proto-clusters in Emission: A Large-scale Structure at z = 2.44 and the Outlook for HETDEX", ApJ, vol. 808, no. 1, 37, 2015. 1505.03877, URL http://dx.doi.org/10.1088/0004-637X/808/1/37.

Clay, S. J., Thomas, P. A., Wilkins, S. M. and Henriques, B. M. B., "Galaxy formation in the Planck cosmology - III. The high-redshift universe", MNRAS, vol. 451, no. 3, pp. 2692–2702, 2015. 1504.03321, URL http://dx.doi.org/10.1093/mnras/stv818.

Cole, S., Lacey, C. G., Baugh, C. M. and Frenk, C. S., "Hierarchical galaxy formation", MNRAS, vol. 319, no. 1, pp. 168–204, 2000. astro-ph/0007281, URL http://dx.doi.org/10.1046/j.1365-8711.2000.03879.x.

Colless, M. et al., "The 2dF Galaxy Redshift Survey: spectra and redshifts", MNRAS, vol. 328, no. 4, pp. 1039–1063, 2001. astro-ph/0106498, URL http://dx.doi.org/10.1046/j.1365-8711.2001.04902.x.

Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E. and Kurk, J. D., "A z = 2.5 protocluster associated with the radio galaxy MRC 2104-242: star formation and differing mass functions in dense environments", MNRAS, vol. 440, no. 4, pp. 3262–3274, 2014. 1403.4259, URL http://dx.doi.org/10.1093/mnras/stu522.

Cooke, E. A. et al., "A Mature Galaxy Cluster at z=1.58 around the Radio Galaxy 7C1753+6311", ApJ, vol. 816, no. 2, 83, 2016. 1511.05150, URL http://dx.doi.org/10.3847/0004-637X/816/2/83.

Costa-Duarte, M. V. et al., "The galaxy environment in GAMA G3C groups using the Kilo Degree Survey Data Release 3", MNRAS, vol. 478, no. 2, pp. 1968–1979, 2018. 1712.07670, URL http://dx.doi.org/10.1093/mnras/sty944.

Coupon, J. et al., "Photometric redshifts for the CFHTLS T0004 deep and wide fields", A&A, vol. 500, no. 3, pp. 981–998, 2009. 0811.3326, URL http://dx.doi.org/10.1051/0004-6361/200811413.

Cucciati, O. et al., "The progeny of a cosmic titan: a massive multi-component proto-supercluster in formation at z = 2.45 in VUDS", A&A, vol. 619, A49, 2018. 1806.06073, URL http://dx.doi.org/10.1051/0004-6361/201833655.

Dale, D. A. et al., "A Two-parameter Model for the Infrared/Submillimeter/Radio Spectral Energy Distributions of Galaxies and Active Galactic Nuclei", ApJ, vol. 784, no. 1, 83, 2014. 1402.1495, URL http://dx.doi.org/10.1088/0004-637X/784/1/83.

Davis, M., Efstathiou, G., Frenk, C. S. and White, S. D. M., "The evolution of large-scale structure in a universe dominated by cold dark matter", ApJ, vol. 292, pp. 371–394, 1985. URL http://dx.doi.org/10.1086/163168.

De Lucia, G. and Blaizot, J., "The hierarchical formation of the brightest cluster galaxies", MNRAS, vol. 375, no. 1, pp. 2–14, 2007. astro-ph/0606519, URL http://dx.doi.org/10.1111/j.1365-2966.2006.11287.x.

Dey, A. et al., "Spectroscopic Confirmation of a Protocluster at z ≈ 3.786", ApJ, vol. 823, no. 1, 11, 2016. 1604.08627, URL http://dx.doi.org/10.3847/0004-637X/823/1/11.

Dietrich, J. P. et al., "Orientation bias of optically selected galaxy clusters and its impact on stacked weak-lensing analyses", MNRAS, vol. 443, no. 2, pp. 1713–1722, 2014. 1405.2923, URL http://dx.doi.org/10.1093/mnras/stu1282.

Draine, B. T. and Li, A., "Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era", ApJ, vol. 657, no. 2, pp. 810–837, 2007. astro-ph/0608003, URL http://dx.doi.org/10.1086/511055.

Dressler, A., "Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies.", ApJ, vol. 236, pp. 351–365, 1980. URL http://dx.doi.org/10.1086/157753.

Dubois, Y. et al., "Dancing in the dark: galactic properties trace spin swings along the cosmic web", MNRAS, vol. 444, no. 2, pp. 1453–1468, 2014. 1402.1165, URL http://dx.doi.org/10.1093/mnras/stu1227.

Durret, F. et al., "Galaxy cluster searches based on photometric redshifts in the four CFHTLS Wide fields", A&A, vol. 535, A65, 2011. 1109.0850, URL http://dx.doi.org/10.1051/0004-6361/201116985.

Ebeling, H., Edge, A. C. and Henry, J. P., "MACS: A Quest for the Most Massive Galaxy Clusters in the Universe", ApJ, vol. 553, no. 2, pp. 668–676, 2001. astro-ph/0009101, URL http://dx.doi.org/10.1086/320958.

Ebeling, H. et al., "The ROSAT Brightest Cluster Sample - IV. The extended sample", MNRAS, vol. 318, no. 2, pp. 333–340, 2000. astro-ph/0003191, URL http://dx.doi.org/10.1046/j.1365-8711.2000.03549.x.

Euclid Collaboration et al., "Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection", A&A, vol. 627, A23, 2019. 1906.04707, URL http://dx.doi.org/10.1051/0004-6361/201935088.

Fitzpatrick, E. L. and Massa, D., "An Analysis of the Shapes of Ultraviolet Extinction Curves. I. The 2175 Angstrom Bump", ApJ, vol. 307, p. 286, 1986. URL http://dx.doi.org/10.1086/164415.

Gladders, M. D., López-Cruz, O., Yee, H. K. C. and Kodama, T., "The Slope of the Cluster Elliptical Red Sequence: A Probe of Cluster Evolution", ApJ, vol. 501, no. 2, pp. 571–577, 1998. astro-ph/9802167, URL http://dx.doi.org/10.1086/305858.

Gladders, M. D. and Yee, H. K. C., "The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29", ApJS, vol. 157, no. 1, pp. 1–29, 2005. astro-ph/0411075, URL http://dx.doi.org/10.1086/427327.

Gonzalez, A., "Cluster detection via wavelets", in: *Building the Euclid Cluster Survey - Scientific Program*, p. 7 (2014).

Gonzalez, A. H. et al., "The Massive and Distant Clusters of WISE Survey. I. Survey Overview and a Catalog of >2000 Galaxy Clusters at z ∼ 1", ApJS, vol. 240, no. 2, 33, 2019. 1809.06820, URL http://dx.doi.org/10.3847/1538-4365/aafad2.

Graham, M. L. et al., "Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies", AJ, vol. 155, no. 1, 1, 2018. 1706.09507, URL http://dx.doi.org/10.3847/1538-3881/aa99d4.

Graham, M. L. et al., "Photometric Redshifts with the LSST II: The Impact of Near-Infrared and Near-Ultraviolet Photometry", *arXiv e-prints*, arXiv:2004.07885, 2020. 2004.07885.

Guiderdoni, B. and Rocca-Volmerange, B., "A model of spectrophotometric evolution for high-redshift galaxies.", A&A, vol. 186, pp. 1–21, 1987.

Gunn, J. E. and Gott, I., J. Richard, "On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution", ApJ, vol. 176, p. 1, 1972. URL http://dx.doi.org/10.1086/151605.

Hamuy, M. et al., "The Absolute Luminosities of the Calan/Tololo Type IA Supernovae", AJ, vol. 112, p. 2391, 1996. astro-ph/9609059, URL http://dx.doi.org/10.1086/118190.

Harrison, C. M., Meiksin, A. and Stock, D., "IGMtransmission: A Java GUI to model the effects of the Intergalactic Medium on the colours of high redshift galaxies", *arXiv e-prints*, arXiv:1105.6208, 2011. 1105.6208.

Hatch, N. A. et al., "Galaxy protocluster candidates around z∼ 2.4 radio galaxies", MNRAS, vol. 410, no. 3, pp. 1537–1549, 2011a. 1008.4588, URL http://dx.doi.org/10.1111/j.1365-2966.2010.17538.x.

Hatch, N. A. et al., "Hα emitters in z∼ 2 protoclusters: evidence for faster evolution in dense environments", MNRAS, vol. 415, no. 4, pp. 2993–3005, 2011b. 1103.4364, URL http://dx.doi.org/10.1111/j.1365-2966.2011.18735.x.

Hatch, N. A. et al., "Why $z > 1$ radio-loud galaxies are commonly located in protoclusters", MNRAS, vol. 445, no. 1, pp. 280–289, 2014. 1409.1218, URL http://dx.doi.org/10.1093/mnras/stu1725.

Hatch, N. A. et al., "The impact of protocluster environments at z = 1.6", MNRAS, vol. 464, no. 1, pp. 876–884, 2017. 1609.08629, URL http://dx.doi.org/10.1093/mnras/stw2359.

Hayashi, M., Kodama, T., Tadaki, K.-i., Koyama, Y. and Tanaka, I., "A Starbursting Protocluster in Making Associated with a Radio Galaxy at z = 2.53 Discovered by Hα Imaging", ApJ, vol. 757, no. 1, 15, 2012. 1207.2614, URL http://dx.doi.org/10.1088/0004-637X/757/1/15.

Henriques, B. M. B. et al., "Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses", MNRAS, vol. 451, no. 3, pp. 2663–2680, 2015. 1410.0365, URL http://dx.doi.org/10.1093/mnras/stv705.

Higuchi, R. et al., "SILVERRUSH. VII. Subaru/HSC Identifications of Protocluster Candidates at z ∼ 6–7: Implications for Cosmic Reionization", ApJ, vol. 879, no. 1, 28, 2019. 1801.00531, URL http://dx.doi.org/10.3847/1538-4357/ab2192.

Hou, A., Parker, L. C. and Harris, W. E., "The pre-processing of subhaloes in SDSS groups and clusters", MNRAS, vol. 442, no. 1, pp. 406–418, 2014. 1404.7504, URL http://dx.doi.org/10.1093/mnras/stu829.

Ilbert, O. et al., "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey", A&A, vol. 457, no. 3, pp. 841–856, 2006. astro-ph/0603217, URL http://dx.doi.org/10.1051/0004-6361:20065138.

Ilbert, O. et al., "Cosmos Photometric Redshifts with 30-Bands for 2-deg^2", ApJ, vol. 690, no. 2, pp. 1236–1249, 2009. 0809.2101, URL http://dx.doi.org/10.1088/0004-637X/690/2/1236.

Ilbert, O. et al., "Mass assembly in quiescent and star-forming galaxies since z ∼ 4 from UltraVISTA", A&A, vol. 556, A55, 2013. 1301.3157, URL http://dx.doi.org/10.1051/0004-6361/201321100.

Inoue, A. K. and Iwata, I., "A Monte Carlo simulation of the intergalactic absorption and the detectability of the Lyman continuum from distant galaxies", MNRAS, vol. 387, no. 4, pp. 1681–1692, 2008. 0804.2951, URL http://dx.doi.org/10.1111/j.1365-2966.2008.13350.x.

Ivezic, Z. et al., "Large Synoptic Survey Telescope: From Science Drivers To Reference Design", *Serbian Astronomical Journal*, vol. 176, pp. 1–13, 2008. URL http://dx.doi.org/10.2298/SAJ0876001I.

Ivezić, Ž. et al., "LSST: From Science Drivers to Reference Design and Anticipated Data Products", ApJ, vol. 873, no. 2, 111, 2019. 0805.2366, URL http://dx.doi.org/10.3847/1538-4357/ab042c.

Izquierdo-Villalba, D. et al., "J-PLUS: Synthetic galaxy catalogues with emission lines for photometric surveys", A&A, vol. 631, A82, 2019. 1907.02111, URL http://dx.doi.org/10.1051/0004-6361/201936232.

Jenkins, A. et al., "The mass function of dark matter haloes", MNRAS, vol. 321, no. 2, pp. 372–384, 2001. astro-ph/0005260, URL http://dx.doi.org/10.1046/j.1365-8711.2001.04029.x.

Jiang, L. et al., "A giant protocluster of galaxies at redshift 5.7", *Nature Astronomy*, vol. 2, pp. 962–966, 2018. 1810.05765, URL http://dx.doi.org/10.1038/s41550-018-0587-9.

Kashikawa, N. et al., "The Subaru Deep Field: The Optical Imaging Data", PASJ, vol. 56, pp. 1011–1023, 2004. astro-ph/0410005, URL http://dx.doi.org/10.1093/pasj/56.6.1011.

Kim, J.-W. et al., "Discovery of a Supercluster at z ~0.91 and Testing the ΛCDM Cosmological Model", ApJ, vol. 821, no. 1, L10, 2016. 1604.03254, URL http://dx.doi.org/10.3847/2041-8205/821/1/L10.

Kitzbichler, M. G. and White, S. D. M., "The high-redshift galaxy population in hierarchical galaxy formation models", MNRAS, vol. 376, no. 1, pp. 2–12, 2007. astro-ph/0609636, URL http://dx.doi.org/10.1111/j.1365-2966.2007.11458.x.

Klypin, A. A., Trujillo-Gomez, S. and Primack, J., "Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation", ApJ, vol. 740, no. 2, 102, 2011. 1002.3660, URL http://dx.doi.org/10.1088/0004-637X/740/2/102.

Kolb, E. W., "Particle Physics in the Early Universe", *arXiv e-prints*, hep-ph/9810362, 1998. hep-ph/9810362.

Koyama, Y. et al., "Massive starburst galaxies in a z = 2.16 proto-cluster unveiled by panoramic Hα mapping", MNRAS, vol. 428, no. 2, pp. 1551–1564, 2013. 1210.0972, URL http://dx.doi.org/10.1093/mnras/sts133.

Krefting, N. et al., "The Role of Environment in Galaxy Evolution in the SERVS Survey. I. Density Maps and Cluster Candidates", ApJ, vol. 889, no. 2, 185, 2020. 1912.02238, URL http://dx.doi.org/10.3847/1538-4357/ab60a0.

Laigle, C. et al., "The COSMOS2015 Catalog: Exploring the $1 < z < 6$ Universe with Half a Million Galaxies", ApJS, vol. 224, no. 2, 24, 2016. 1604.02350, URL http://dx.doi.org/10.3847/0067-0049/224/2/24.

Laigle, C. et al., "Horizon-AGN virtual observatory - 1. SED-fitting performance and forecasts for future imaging surveys", MNRAS, vol. 486, no. 4, pp. 5104–5123, 2019. 1903.10934, URL http://dx.doi.org/10.1093/mnras/stz1054.

Le Fèvre, O. et al., "The VIMOS Ultra-Deep Survey: $\sim$10 000 galaxies with spectroscopic redshifts to study galaxy assembly at early epochs $2 < z \simeq 6$", A&A, vol. 576, A79, 2015. 1403.3938, URL http://dx.doi.org/10.1051/0004-6361/201423829.

Leitherer, C. and Heckman, T. M., "Synthetic Properties of Starburst Galaxies", ApJS, vol. 96, p. 9, 1995. URL http://dx.doi.org/10.1086/192112.

Lemaux, B. C. et al., "The VIMOS Ultra-Deep Survey: Emerging from the dark, a massive proto-cluster at z 4.57", A&A, vol. 615, A77, 2018. 1703.10170, URL http://dx.doi.org/10.1051/0004-6361/201730870.

Lilly, S. J. et al., "zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering 0 < z < 3 in the COSMOS Field", ApJS, vol. 172, no. 1, pp. 70–85, 2007. astro-ph/0612291, URL http://dx.doi.org/10.1086/516589.

Lilly, S. J. et al., "The zCOSMOS 10k-Bright Spectroscopic Sample", ApJS, vol. 184, no. 2, pp. 218–229, 2009. URL http://dx.doi.org/10.1088/0067-0049/184/2/218.

Lima, E. V. R. et al., "Photometric redshifts for S-PLUS using machine learning techniques", , in prep.

Lima Neto, G. B., Lagana, T. F., Andrade-Santos, F. and Machado, R. E. G., "Structure in galaxy clusters", *arXiv e-prints*, arXiv:1406.1496, 2014. 1406.1496.

Lin, Y.-T. and Mohr, J. J., "K-band Properties of Galaxy Clusters and Groups: Brightest Cluster Galaxies and Intracluster Light", ApJ, vol. 617, no. 2, pp. 879–895, 2004. astro-ph/0408557, URL http://dx.doi.org/10.1086/425412.

Long, A. S. et al., "Emergence of an Ultra-Red Ultra-Massive Galaxy Cluster Core at $z = 4$", *arXiv e-prints*, arXiv:2003.13694, 2020. 2003.13694.

Lovell, C. C., Thomas, P. A. and Wilkins, S. M., "Characterising and identifying galaxy protoclusters", MNRAS, vol. 474, no. 4, pp. 4612–4628, 2018. 1710.02148, URL http://dx.doi.org/10.1093/mnras/stx3090.

Maraston, C., "Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies", MNRAS, vol. 362, no. 3, pp. 799–825, 2005. astro-ph/0410207, URL http://dx.doi.org/10.1111/j.1365-2966.2005.09270.x.

Martinache, C. et al., "Spitzer Planck Herschel Infrared Cluster (SPHerIC) survey: Candidate galaxy clusters at 1.3 < z < 3 selected by high star-formation rate", A&A, vol. 620, A198, 2018. 1810.07330, URL http://dx.doi.org/10.1051/0004-6361/201833198.

Mathis, J. S., Mezger, P. G. and Panagia, N., "Interstellar radiation field and dust temperatures in the diffuse interstellar matter and in giant molecular clouds.", A&A, vol. 500, pp. 259–276, 1983.

McCracken, H. J. et al., "The VIRMOS deep imaging survey. II: CFH12K BVRI optical data for the 0226-04 deep field", A&A, vol. 410, pp. 17–32, 2003. astro-ph/0306254, URL http://dx.doi.org/10.1051/0004-6361:20031081.

Merson, A. I. et al., "Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection", MNRAS, vol. 429, no. 1, pp. 556–578, 2013. 1206.4049, URL http://dx.doi.org/10.1093/mnras/sts355.

Metcalfe, N., Shanks, T., Campos, A., McCracken, H. J. and Fong, R., "Galaxy number counts - V. Ultradeep counts: the Herschel and Hubble Deep Fields", MNRAS, vol. 323, no. 4, pp. 795–830, 2001. astro-ph/0010153, URL http://dx.doi.org/10.1046/j.1365-8711.2001.04168.x.

Miller, T. B. et al., "A massive core for a cluster of galaxies at a redshift of 4.3", Nature, vol. 556, no. 7702, pp. 469–472, 2018. 1804.09231, URL http://dx.doi.org/10.1038/s41586-018-0025-2.

Molino, A. et al., "Assessing the photometric redshift precision of the S-PLUS survey: the Stripe-82 as a test-case", *arXiv e-prints*, arXiv:1907.06315, 2019a. 1907.06315.

Molino, A. et al., "J-PLUS: On the identification of new cluster members in the double galaxy cluster A2589 and A2593 using PDFs", A&A, vol. 622, A178, 2019b. 1804.03640, URL http://dx.doi.org/10.1051/0004-6361/201731348.

Moore, B., Lake, G. and Katz, N., "Morphological Transformation from Galaxy Harassment", ApJ, vol. 495, no. 1, pp. 139–151, 1998. astro-ph/9701211, URL http://dx.doi.org/10.1086/305264.

Muldrew, S. I., Hatch, N. A. and Cooke, E. A., "What are protoclusters? - Defining high-redshift galaxy clusters and protoclusters", MNRAS, vol. 452, no. 3, pp. 2528–2539, 2015. 1506.08835, URL http://dx.doi.org/10.1093/mnras/stv1449.

Muldrew, S. I., Hatch, N. A. and Cooke, E. A., "Galaxy evolution in protoclusters", MNRAS, vol. 473, no. 2, pp. 2335–2347, 2018. 1709.07009, URL http://dx.doi.org/10.1093/mnras/stx2454.

Muzzin, A. et al., "A Public K_s -selected Catalog in the COSMOS/ULTRAVISTA Field: Photometry, Photometric Redshifts, and Stellar Population Parameters", ApJS, vol. 206, no. 1, 8, 2013. 1303.4410, URL http://dx.doi.org/10.1088/0067-0049/206/1/8.

Oke, J. B. and Gunn, J. E., "Secondary standard stars for absolute spectrophotometry.", ApJ, vol. 266, pp. 713–717, 1983. URL http://dx.doi.org/10.1086/160817.

O'Mill, A. L., Duplancic, F., García Lambas, D. and Sodré, J., Laerte, "Photometric redshifts and k-corrections for the Sloan Digital Sky Survey Data Release 7", MNRAS, vol. 413, no. 2, pp. 1395–1408, 2011. 1012.3752, URL http://dx.doi.org/10.1111/j.1365-2966.2011.18222.x.

Onoue, M. et al., "Enhancement of galaxy overdensity around quasar pairs at z < 3.6 based on the Hyper Suprime-Cam Subaru Strategic Program Survey", PASJ, vol. 70, S31, 2018. 1704.06051, URL http://dx.doi.org/10.1093/pasj/psx092.

Orsi, Á. et al., "The nebular emission of star-forming galaxies in a hierarchical universe", MNRAS, vol. 443, no. 1, pp. 799–814, 2014. 1402.5145, URL http://dx.doi.org/10.1093/mnras/stu1203.

Orsi, Á. A., Fanidakis, N., Lacey, C. G. and Baugh, C. M., "The environments of high-redshift radio galaxies and quasars: probes of protoclusters", MNRAS, vol. 456, no. 4, pp. 3827–3839, 2016. 1509.01254, URL http://dx.doi.org/10.1093/mnras/stv2919.

Overzier, R. et al., "The Millennium Run Observatory: first light", MNRAS, vol. 428, no. 1, pp. 778–803, 2013. 1206.6923, URL http://dx.doi.org/10.1093/mnras/sts076.

Overzier, R. A., "The realm of the galaxy protoclusters. A review", A&A Rev., vol. 24, no. 1, 14, 2016. 1610.05201, URL http://dx.doi.org/10.1007/s00159-016-0100-3.

Overzier, R. A. et al., "Clustering of Star-forming Galaxies Near a Radio Galaxy at z=5.2", ApJ, vol. 637, no. 1, pp. 58–73, 2006. astro-ph/0509308, URL http://dx.doi.org/10.1086/498234.

Overzier, R. A. et al., "Lyman Break Galaxies, Lyα Emitters, and a Radio Galaxy in a Protocluster at z = 4.1", ApJ, vol. 673, no. 1, pp. 143–162, 2008. astro-ph/0601223, URL http://dx.doi.org/10.1086/524342.

Overzier, R. A. et al., "Stellar Masses of Lyman Break Galaxies, Lyα Emitters, and Radio Galaxies in Overdense Regions at z = 4-6", ApJ, vol. 704, no. 1, pp. 548–563, 2009. 0909.1082, URL http://dx.doi.org/10.1088/0004-637X/704/1/548.

Pallero, D. et al., "Tracing the quenching history of cluster galaxies in the EAGLE simulation", MNRAS, vol. 488, no. 1, pp. 847–858, 2019. 1812.08802, URL http://dx.doi.org/10.1093/mnras/stz1745.

Pentericci, L. et al., "A search for clusters at high redshift. II. A proto cluster around a radio galaxy at z=2.16", A&A, vol. 361, pp. L25–L28, 2000. astro-ph/0008143.

Perlmutter, S. et al., "Measurements of Ω and Λ from 42 High-Redshift Supernovae", ApJ, vol. 517, no. 2, pp. 565–586, 1999. astro-ph/9812133, URL http://dx.doi.org/10.1086/307221.

Phillips, M. M., "The Absolute Magnitudes of Type IA Supernovae", ApJ, vol. 413, p. L105, 1993. URL http://dx.doi.org/10.1086/186970.

Planck Collaboration et al., "Planck 2013 results. XVI. Cosmological parameters", A&A, vol. 571, A16, 2014. 1303.5076, URL http://dx.doi.org/10.1051/0004-6361/201321591.

Planck Collaboration et al., "Planck 2015 results. XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources", A&A, vol. 594, A27, 2016. 1502.01598, URL http://dx.doi.org/10.1051/0004-6361/201525823.

Polletta, M. et al., "Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey", ApJ, vol. 663, no. 1, pp. 81–102, 2007. astro-ph/0703255, URL http://dx.doi.org/10.1086/518113.

Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E. and Primack, J., "Halo concentrations in the standard Λ cold dark matter cosmology", MNRAS, vol. 423, no. 4, pp. 3018–3030, 2012. 1104.5130, URL http://dx.doi.org/10.1111/j.1365-2966.2012.21007.x.

Press, W. H. and Schechter, P., "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation", ApJ, vol. 187, pp. 425–438, 1974. URL http://dx.doi.org/10.1086/152650.

Prevot, M. L., Lequeux, J., Maurice, E., Prevot, L. and Rocca-Volmerange, B., "The typical interstellar extinction in the Small Magellanic Cloud.", A&A, vol. 132, pp. 389–392, 1984.

Riess, A. G. et al., "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant", AJ, vol. 116, no. 3, pp. 1009–1038, 1998. astro-ph/9805201, URL http://dx.doi.org/10.1086/300499.

Rotermund, K. M. et al., "Optical and near-infrared observations of the SPT2349-56 proto-cluster core at $z = 4.3$", *arXiv e-prints*, arXiv:2006.15345, 2020. 2006.15345.

Rovilos, E. et al., "Deep U-B-V imaging of the Lockman Hole with the LBT. Observations and number counts", A&A, vol. 507, no. 1, pp. 195–208, 2009. 0909.0661, URL http://dx.doi.org/10.1051/0004-6361/200912626.

Rykoff, E. S. et al., "redMaPPer. I. Algorithm and SDSS DR8 Catalog", ApJ, vol. 785, no. 2, 104, 2014. 1303.3562, URL http://dx.doi.org/10.1088/0004-637X/785/2/104.

Sadeh, I., Abdalla, F. B. and Lahav, O., "ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning", PASP, vol. 128, no. 968, p. 104502, 2016. 1507.00490, URL http://dx.doi.org/10.1088/1538-3873/128/968/104502.

Sánchez, C. et al., "Photometric redshift analysis in the Dark Energy Survey Science Verification data", MNRAS, vol. 445, no. 2, pp. 1482–1506, 2014. 1406.4407, URL http://dx.doi.org/10.1093/mnras/stu1836.

Santos, J. S. et al., "Discovery of a massive X-ray luminous galaxy cluster at z = 1.579", A&A, vol. 531, L15, 2011. 1105.5877, URL http://dx.doi.org/10.1051/0004-6361/201117190.

Schaye, J. et al., "The EAGLE project: simulating the evolution and assembly of galaxies and their environments", MNRAS, vol. 446, no. 1, pp. 521–554, 2015. 1407.7040, URL http://dx.doi.org/10.1093/mnras/stu2058.

Schneider, P., *Extragalactic Astronomy and Cosmology* (Springer-Verlag Berlin Heidelberg, 2006).

Shamshiri, S. et al., "Galaxy formation in the Planck cosmology - II. Star-formation histories and post-processing magnitude reconstruction", MNRAS, vol. 451, no. 3, pp. 2681–2691, 2015. 1501.05649, URL http://dx.doi.org/10.1093/mnras/stv883.

Shi, K. et al., "A Census of Galaxy Constituents in a Coma Progenitor Observed at z > 3", ApJ, vol. 871, no. 1, 83, 2019. 1808.07065, URL http://dx.doi.org/10.3847/1538-4357/aaf85d.

Shimakawa, R. et al., "MAHALO Deep Cluster Survey I. Accelerated and enhanced galaxy formation in the densest regions of a protocluster at z = 2.5", MNRAS, vol. 473, no. 2, pp. 1977–1999, 2018a. 1708.06369, URL http://dx.doi.org/10.1093/mnras/stx2494.

Shimakawa, R. et al., "MAHALO Deep Cluster Survey II. Characterizing massive forming galaxies in the Spiderweb protocluster at z = 2.2", MNRAS, vol. 481, no. 4, pp. 5630–5650, 2018b. 1809.08755, URL http://dx.doi.org/10.1093/mnras/sty2618.

Spergel, D. N. et al., "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters", ApJS, vol. 148, no. 1, pp. 175–194, 2003. astro-ph/0302209, URL http://dx.doi.org/10.1086/377226.

Springel, V., "The cosmological simulation code GADGET-2", MNRAS, vol. 364, no. 4, pp. 1105–1134, 2005. astro-ph/0505010, URL http://dx.doi.org/10.1111/j.1365-2966.2005.09655.x.

Springel, V., Frenk, C. S. and White, S. D. M., "The large-scale structure of the Universe", Nature, vol. 440, no. 7088, pp. 1137–1144, 2006. astro-ph/0604561, URL http://dx.doi.org/10.1038/nature04805.

Springel, V., White, S. D. M., Tormen, G. and Kauffmann, G., "Populating a cluster of galaxies - I. Results at [formmu2]z=0", MNRAS, vol. 328, no. 3, pp. 726–750, 2001. astro-ph/0012055, URL http://dx.doi.org/10.1046/j.1365-8711.2001.04912.x.

Springel, V. et al., "Simulations of the formation, evolution and clustering of galaxies and quasars", Nature, vol. 435, no. 7042, pp. 629–636, 2005. astro-ph/0504097, URL http://dx.doi.org/10.1038/nature03597.

Staniszewski, Z. et al., "Galaxy Clusters Discovered with a Sunyaev-Zel'dovich Effect Survey", ApJ, vol. 701, no. 1, pp. 32–41, 2009. 0810.1578, URL http://dx.doi.org/10.1088/0004-637X/701/1/32.

Steidel, C. C. et al., "A Large Structure of Galaxies at Redshift Z approximately 3 and Its Cosmological Implications", ApJ, vol. 492, no. 2, pp. 428–438, 1998. astro-ph/9708125, URL http://dx.doi.org/10.1086/305073.

Stothert, L. et al., "The PAU Survey: spectral features and galaxy clustering using simulated narrow-band photometry", MNRAS, vol. 481, no. 3, pp. 4221–4235, 2018. 1807.03260, URL http://dx.doi.org/10.1093/mnras/sty2491.

Stott, J. P. et al., "Quasar Sightline and Galaxy Evolution (QSAGE) survey – II. Galaxy overdensities around UV luminous quasars at z=1-2", *arXiv e-prints*, arXiv:2006.07384, 2020. 2006.07384.

Strazzullo, V. et al., "Galaxy populations in the most distant SPT-SZ clusters. I. Environmental quenching in massive clusters at $1.4 \lesssim z \lesssim 1.7$", A&A, vol. 622, A117, 2019. 1807.09768, URL http://dx.doi.org/10.1051/0004-6361/201833944.

Suzuki, N. et al., "The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above $z > 1$ and Building an Early-type-hosted Supernova Sample", ApJ, vol. 746, no. 1, 85, 2012. 1105.3470, URL http://dx.doi.org/10.1088/0004-637X/746/1/85.

Takada, M. et al., "Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph", PASJ, vol. 66, no. 1, R1, 2014. 1206.0737, URL http://dx.doi.org/10.1093/pasj/pst019.

Tanaka, M. et al., "Photometric redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1", PASJ, vol. 70, S9, 2018. 1704.05988, URL http://dx.doi.org/10.1093/pasj/psx077.

Toshikawa, J. et al., "A Systematic Survey of Protoclusters at z ~3-6 in the CFHTLS Deep Fields", ApJ, vol. 826, no. 2, 114, 2016. 1605.01439, URL http://dx.doi.org/10.3847/0004-637X/826/2/114.

Toshikawa, J. et al., "GOLDRUSH. III. A systematic search for protoclusters at z $\sim$ 4 based on the $> 100 deg^2$ area", PASJ, vol. 70, S12, 2018. 1708.09421, URL http://dx.doi.org/10.1093/pasj/psx102.

Trebitsch, M. et al., "The Obelisk simulation: galaxies contribute more than AGN to HI reionization of protoclusters", *arXiv e-prints*, arXiv:2002.04045, 2020. 2002.04045.

Uchiyama, H. et al., "Luminous quasars do not live in the most overdense regions of galaxies at z $\sim$ 4", PASJ, vol. 70, S32, 2018. 1704.06050, URL http://dx.doi.org/10.1093/pasj/psx112.

Venemans, B. P. et al., "Properties of Lyα emitters around the radio galaxy MRC 0316 257'", A&A, vol. 431, no. 3, pp. 793–812, 2005. astro-ph/0501259, URL http://dx.doi.org/10.1051/0004-6361:20042038.

Venemans, B. P. et al., "Protoclusters associated with z > 2 radio galaxies . I. Characteristics of high redshift protoclusters", A&A, vol. 461, no. 3, pp. 823–845, 2007. astro-ph/0610567, URL http://dx.doi.org/10.1051/0004-6361:20053941.

Vicentin, M. C. et al., "The environment of QSO triplets at $1 \lesssim z \lesssim 1.5$", , submitted.

Vikhlinin, A. A., Kravtsov, A. V., Markevich, M. L., Sunyaev, R. A. and Churazov, E. M., "Clusters of galaxies", *Physics Uspekhi*, vol. 57, no. 4, 317-341, 2014. URL http://dx.doi.org/10.3367/UFNe.0184.201404a.0339.

Vogelsberger, M. et al., "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe", MNRAS, vol. 444, no. 2, pp. 1518–1547, 2014. 1405.2921, URL http://dx.doi.org/10.1093/mnras/stu1536.

Watson, C. et al., "Galaxy Merger Fractions in Two Clusters at $z \sim 2$ Using the Hubble Space Telescope", ApJ, vol. 874, no. 1, 63, 2019. 1902.07225, URL http://dx.doi.org/10.3847/1538-4357/ab06ef.

Wen, Z. L. and Han, J. L., "Galaxy Clusters at High Redshift and Evolution of Brightest Cluster Galaxies", ApJ, vol. 734, no. 1, 68, 2011. 1104.1667, URL http://dx.doi.org/10.1088/0004-637X/734/1/68.

Werner, S. V. et al., in prep.

Wylezalek, D. et al., "Galaxy Clusters around Radio-loud Active Galactic Nuclei at 1.3 < z < 3.2 as Seen by Spitzer", ApJ, vol. 769, no. 1, 79, 2013. 1304.0770, URL http://dx.doi.org/10.1088/0004-637X/769/1/79.

Yasuda, N. et al., "Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data", AJ, vol. 122, no. 3, pp. 1104–1124, 2001. astro-ph/0105545, URL http://dx.doi.org/10.1086/322093.

Yoon, Y. et al., "Extremely Massive Quasars Are Not Good Proxies for Dense Environments Compared to Massive Galaxies: Environments of Extremely Massive Quasars and Galaxies", ApJ, vol. 871, no. 1, 57, 2019. 1901.02267, URL http://dx.doi.org/10.3847/1538-4357/aaf6e3.

York, D. G. et al., "The Sloan Digital Sky Survey: Technical Summary", AJ, vol. 120, no. 3, pp. 1579–1587, 2000. astro-ph/0006396, URL http://dx.doi.org/10.1086/301513.

Zabludoff, A. I. and Mulchaey, J. S., "The Properties of Poor Groups of Galaxies. I. Spectroscopic Survey and Results", ApJ, vol. 496, no. 1, pp. 39–72, 1998. astro-ph/9708132, URL http://dx.doi.org/10.1086/305355.

A   Appendix

A.1 │ Modeling the overdensity of protoclusters

We have estimated the probability of a given peak with overdensity, δ_{gal}, be a real structure (Figure 4.5). Notice that we have classified them as peaks in Coma-type (1), Virgo-type (2), Fornax-type (3) protoclusters, and in a non-protocluster region (4). To do this, we have generated random values by sampling overdensity distributions with a Gamma function for the four classes. Our main motivation was to avoid dealing with small samples and the effects of statistical fluctuations. Although we have analyzed 80 PCcones per redshift slab, the number of Coma-type protoclusters is still low. For example, at $z = 1.0$, we found 981, 420, and 31 Fornax, Virgo, and Coma-type protoclusters, respectively, for the *LSST-like* sample. These respective numbers for *HSC-like* are 1677, 694, and 49. We obtained overdensity measurements from the density contrast maps, and we fit this distribution for each class, assuming that the Gamma function describes them. To do this, we used the `gamma.fit` function from `scipy` Python Package, which output the parameters a,b and c of Equation A.1.

$$f(y) = \frac{x^{a-1}e^{-y}}{\int_0^\infty t^{a-1}e^{-t}dt}; \qquad y = \frac{x-b}{c} \tag{A.1}$$

After, by using the task `gamma.rvs(`a`, loc=`b`, scale=`c`, size=`N`)`, we generate N random values from our modeled Gamma function. N has to be large and consistent with the number density of a given class. We define this quantity as $N = 3 \times 10^{11} n_i$, where n_i is the number density of peaks associated with the i-th class. Then, we compare the overdensity distributions from our measurements with those obtained from the modeling in Figures A.1 and A.2, for the *HSC-like* and *LSST-like* samples, respectively. In general, the sampled distributions are in qualitative agreement to those from the mocks. The overdensities distributions of Coma-type protoclusters appear to be the worst case, but this is due to statistical fluctuations when small samples are analyzed.
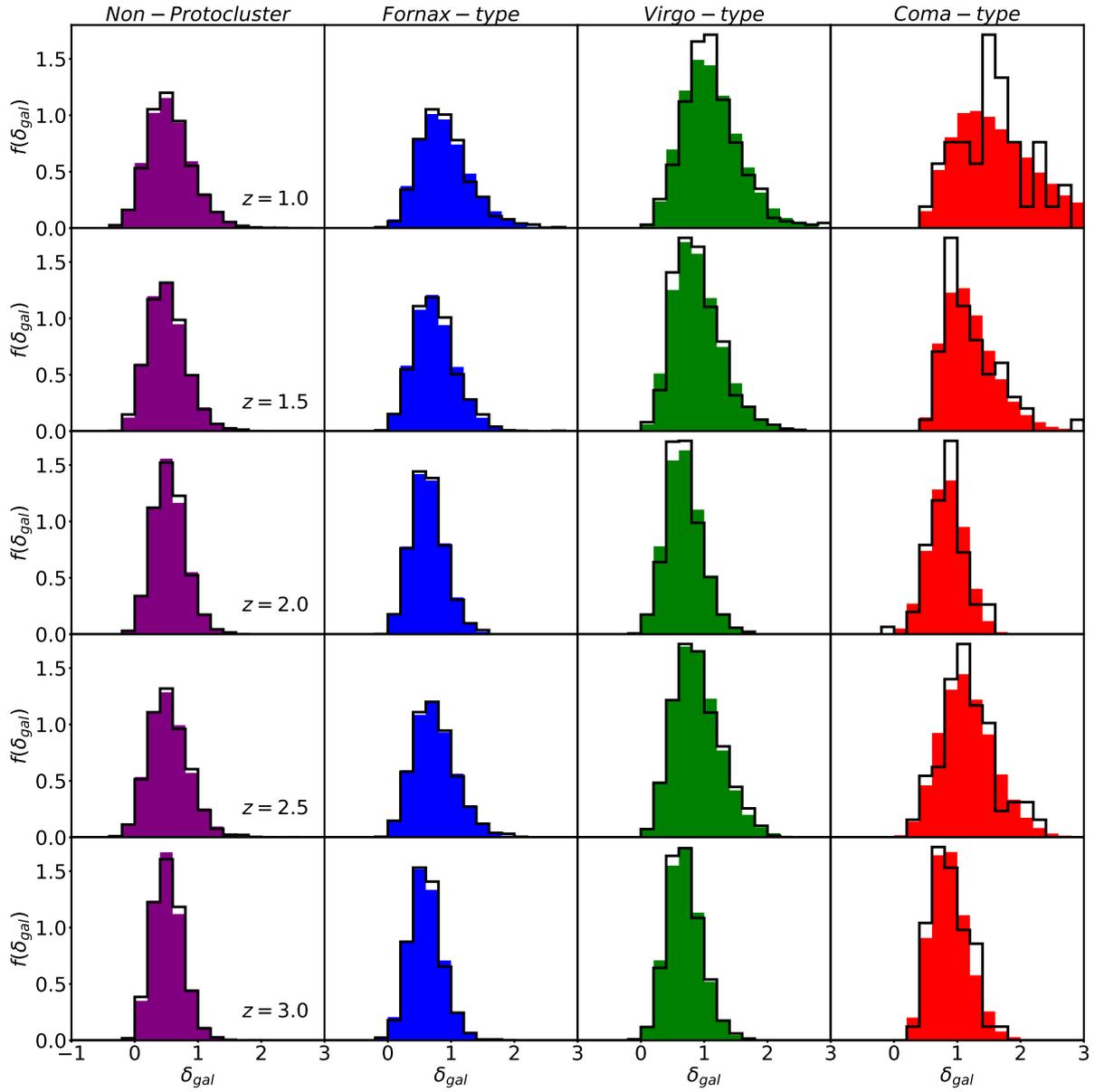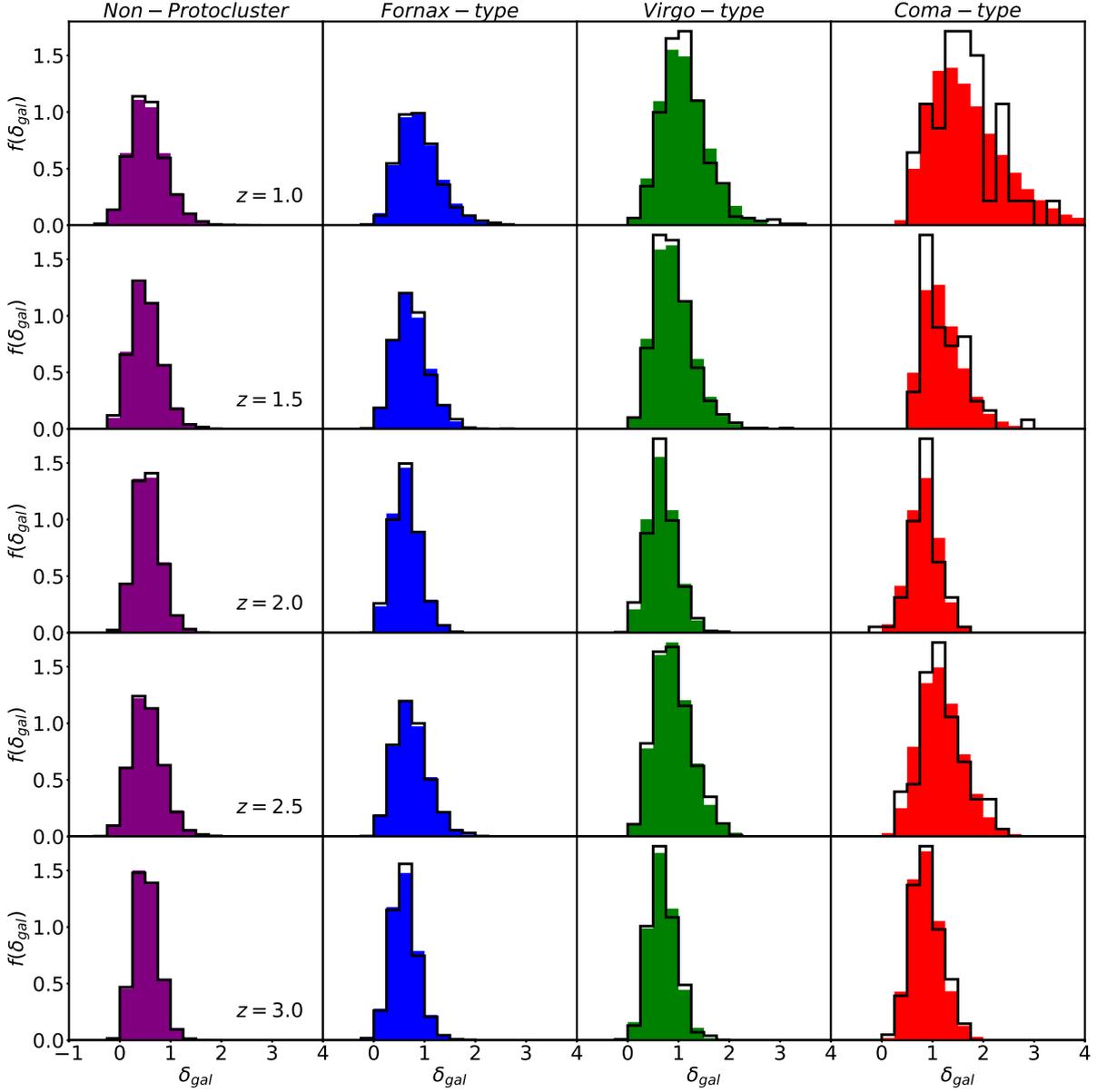
Figure A.1: Probability density distribution of overdensities, for the four different classes: peaks in a non-protocluster region, and in Coma-type, Virgo-type, and Fornax-type protoclusters, from the first to the fourth column respectively. Black solid lines indicates the measurements from the *HSC-like* sample. Filled distributions correspond to the values generated from our model with the Gamma function. Each row corresponds to redshifts $z_p = 1.0$, 1.5, 2.0, 2.5, and 3.0 from top to the bottom.

Figure A.2: Same as Figure A.1 but for the *LSST-like* sample.

Table A.1: *p*-values from the KS test, comparing overdensity distributions from our mock and our Gamma function model. Sub-index C, V, F, and NPC denote the classes Coma, Virgo, Fornax-type protocluster, and peaks in non-protocluster regions, respectively.

	HSC-like				*LSST-like*			
z	$p_{\mathrm{KS},C}$	$p_{\mathrm{KS},V}$	$p_{\mathrm{KS},F}$	$p_{\mathrm{KS},NPC}$	$p_{\mathrm{KS},C}$	$p_{\mathrm{KS},V}$	$p_{\mathrm{KS},F}$	$p_{\mathrm{KS},NPC}$
1.0	0.83	0.50	0.81	0.11	0.84	0.52	0.84	0.11
1.5	0.93	0.78	0.96	0.84	0.97	0.80	0.95	0.89
2.0	0.82	0.35	0.86	0.78	0.80	0.39	0.85	0.78
2.5	0.96	0.88	0.99	0.93	0.94	0.87	0.99	0.91
3.0	0.84	0.77	0.79	0.82	0.86	0.71	0.70	0.77

We quantify the similarity of both overdensity distributions (modeled and from mocks) by computing the *p*-value of the *Kolmogorov-Smirnov* test. We performed this for all class and redshift, where we listed our results in Table A.1. Notice that for all cases, the *p*-values indicate

that both samples are statistically equal.