

The UNIT Simulations: Properties of emission-line galaxies

Daniel López Cano

Máster en Física Teórica



MÁSTERES
DE LA UAM
2019 – 2020

Facultad de Ciencias



FACULTAD DE
CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID



excelencia^{Campus Internacional} UAM
CSIC⁺

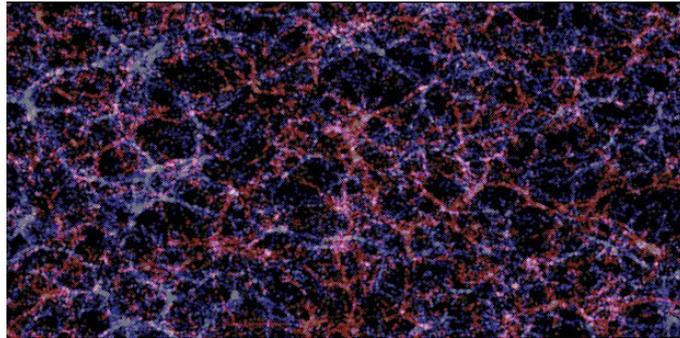
The UNIT Simulations: properties of emission-line galaxies

Universidad Autónoma de Madrid, Department of Theoretical Physics
Master's thesis

Daniel López Cano¹

Supervised by:
Alexander Knebe

Departamento de Física Teórica, Modulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain



Abstract

In order to study some of the most relevant topics within the field of cosmology such as the theory of structure formation, the expansion history of the universe, or the correlation properties associated with the distribution of matter (among many other things), it is necessary to produce catalogues containing the measured positions for as many galaxies as possible. That is why new surveys such as ESA's Euclid mission or NASA's Nancy Grace Roman Space Telescope are planned to start operating over the next years and map with unprecedented precision the 3D positions of tens of millions of galaxies, however, in order to design in the most efficient way the observation plans for missions like these and to estimate their expected performance, it is necessary to develop theoretically modelled galaxy catalogues (typically produced via computational simulations) with which to make predictions and comparisons.

In this work we present four different simulated galaxy catalogues (and explain how we have created them) and once that we have validated the results generated for the simulated galaxies, we provide estimates for the distribution statistics associated with the galaxies that we predict can be observed by the Euclid mission, so that these values can be used in future studies that require the modelling of galaxies in the range of redshifts studied here.

¹ daniellopezcano13@gmail.com

CONTENTS

I. Introduction	3
II. The Methods	5
II.1. The UNIT Simulations	5
II.2. Semi-analytical galaxy modelling via SAGE	5
II.2.1. Cooling	5
II.2.2. Star formation	6
II.2.3. Gas ejection and reincorporation mechanisms	6
II.2.4. Chemical Evolution	6
II.2.5. Merger treatment	7
II.2.6. Calibration method	7
II.3. Emission-line galaxy modelling	7
III. The SAGE galaxies	9
III.1. Stellar Mass Function	9
III.2. Star Formation	10
III.2.1. The Star Formation Rate Function	10
III.2.2. The Specific Star Formation Rate to Stellar Mass Relation	11
III.3. The Mass-Metallicity Relation	12
IV. SAGE's Emission-Line Galaxies (ELGs)	13
IV.1. Abundance of dust-attenuated ELGs	13
IV.2. Abundance evolution of flux-selected ELGs	14
V. Correlation Properties of ELGs	16
VI. Conclusions	19
References	19
Appendices	21
A. Additional validation plots	21
A.1. The Black Hole to Bulge Mass Relation	21
A.2. The Cold Gas Fraction	22
A.3. Stellar-to-Halo Mass Relation	23

I. INTRODUCTION

During the last few decades numerous projects aiming to create large cartographic maps of galaxies such as DES [1], BOSS [21], eBOSS [22], and WiggleZ [26] have been carried out with the objective of trying to better understand the large-scale structure of the universe, to estimate the different parameters that regulate the formation of structures, to attempt to determine the expansion history of the universe, to predict how galaxies form and reconstruct its star formation history, and to impose constraints for the different models that currently exist for dark energy and for alternative theories of gravity different to general relativity.

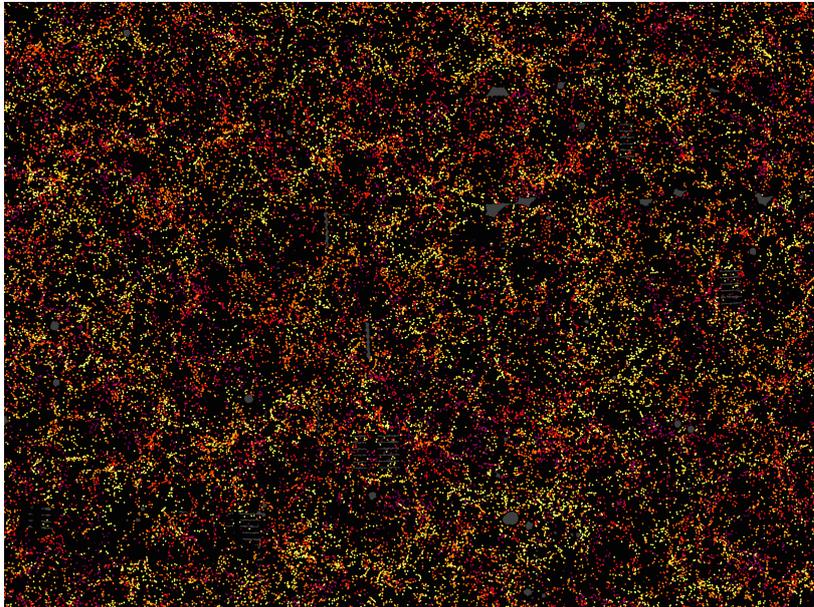


Figure 1: Slice through the map of the large-scale structure of the Universe from the Baryon Oscillation Spectroscopic Survey (BOSS). Each dot in this image represents the position of a galaxy and the color that each one has, its measured redshift (up to $z = 0.7$) [27].

However, many of these topics remain open for debate and new surveys such as Euclid [39][4], the Nancy Grace Roman Space Telescope (formerly known as WFIRST) [48, 49], DESI [16] or 4MOST [55] are expected to start operating in the following years with the objective of providing the scientific community with more accurate measurements by studying larger regions of the night sky and/or collecting more information by surveying deeper fields so that the data gathered for considerably large samples of galaxies may be used to impose stronger constraints upon theoretical models and to provide more accurate estimates for some of the aforementioned parameters relevant in cosmology. Some of these forthcoming missions such as Euclid or WFIRST will focus on conducting spectroscopic surveys of galaxies using near-IR grisms in order to determine the positions of large numbers of galaxies by observing some of its most characteristic emission lines such as H_α and [OIII]. The measurement of the intensity of such emission lines will enable the inference of properties for those galaxies such as their SFR since the presence of these lines in the spectrum indicates that within those galaxies exist ionized [HII] regions resulting from the birthing of new stars [30].

One of the main challenges for these new missions is to optimize the way in which the survey is conducted since there exists a trade-off between the area of the sky that the mission will be able to study and the depth with which it can do so. This is why making predictions for the density of galaxies as a function of redshift and flux [54] as well as for the results derived from their measurements is critical for maximizing the impact of this kind of surveys. To be able to make this kind of predictions it is necessary to develop numerical simulations with which to estimate the number of galaxies that these missions would be able to observe based on their particular characteristics. Since the spatial volumes that these sort of surveys seek to study is notoriously large, it is necessary to rely on dark matter only simulations in which galaxies are introduced in post-processing, for example, using the so called semi-analytical models [37]². While there are efforts to push the limits of full physics hydrodynamical simulations to larger and

² <https://www.cosmosim.org/cms/documentation/projects/galaxies>

larger volumes [40], modelling at least the gas physics in a fully self-consistent way, it still remains more feasible to match the volumes that missions like Euclid foresee to explore ($\sim 1\text{Gpc}^3$ [39],[5]) with dark matter only simulations.

This is also the reason why during the last years, alternatives to reproduce the behaviour of simulations on very big volumes have been explored using techniques that allow to reduce the cosmic variance making use of simulations of smaller sizes. For instance, the technique developed by Angulo & Pontzen [6] dramatically reduces the variance arising from the sparse sampling of wavemodes in cosmological simulations. The method uses two simulations which are ‘fixed’ and ‘paired’, i.e. the initial Fourier mode amplitudes are fixed to the ensemble average power spectrum and their phases are shifted by π . This approach has been adopted by the UNIT collaboration³ [14] and we use two such pairs of simulations for our study here. Such pair can be altogether as precise on non-linear scales as an average over 50 traditional simulations and in Chuang et al. [14] they have shown the the original $(1h^{-1}\text{Gpc})^3$ simulation(s) correspond to an effective volume of approximately $(5h^{-1}\text{Gpc})^3$ (i.e. 7 times of the effective survey volume of DESI or Euclid).

In this work we will focus on making predictions for the two-point correlation function (2pcf) and bias that can be expected to be derived from the observations made by the Euclid experiment by measuring the positions of galaxies with H_α emission line intensities higher than 2×10^{-16} erg/s/cm² for the range of redshifts that it is designed to study ($0.9 < z < 1.8$). The two-point correlation function is a basic estimator in cosmology that gives us information about the probability of finding a galaxy at a given distance assuming that we are located in an arbitrary galaxy, or in James Peebles’ words (one of the 2019 Nobel Laureates in Physics for his major contributions to the field of cosmology): ‘Given a random galaxy in a location, the correlation function describes the probability that another galaxy will be found within a given distance’. On the other hand, the bias is also a fundamental parameter because it allows us to infer which is the underlying dark matter distribution in the universe from the positions of the galaxies that we can observe. The reason for this is that since we cannot directly observe the dark matter distribution but we know that galaxies are found within dark matter halos which have formed due to the gravitational influence of dark matter, the galaxies serve as tracers for this dark matter distribution. The relation between the distribution of galaxies and dark matter can be defined with the bias (b) as presented in Eq.12 through the parameter known as density contrast.

$$\delta_{gal}(\mathbf{x}) = b\delta_{DM}(\mathbf{x}) \quad ; \quad \text{with} \quad \delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\langle\rho\rangle} - 1 \quad (1)$$

This is precisely why determining the value for bias is so important in cosmology, since most of the cosmological simulations that are carried out nowadays provide us with information about the distribution of dark matter in the universe, knowing beforehand the value of bias for the type of galaxies that we are interested in can be very useful in order to connect in a rather direct manner the result of these simulations with the distribution of galaxies that can ultimately be observed, hence facilitating the modeling efforts. It is also important to mention that the formula presented in Eq.12 for a constant bias is expected to hold for large scales (where the linear theory of structure formation is valid), at small enough scales, non-linear effects will cause the bias to no longer remain constant.

In order to make predictions for these quantities we have used the aforementioned two fixed-and-paired dark matter only simulations originally presented in Chuang et al. [14] for which we have generated simulated galaxy catalogues (commonly referred also as galaxy mock catalogues) by applying the SAGE [17] semi-analytical galaxy formation model to them. The H_α emission intensity for the simulated galaxies is then calculated using the model of Orsi et al. [44] since the semi-analytical model applied does not directly produce the values for the intensity of the emission lines of the simulated galaxies, but provides parameters from which we can calculate such intensity. These mock emission-line galaxy (ELG) catalogues will then be used to produce predictions for the 2pcf and bias expected for the redshift range of operation of Euclid.

The structure of this work goes as follows: In section II the methods used to generate the ELGs catalogues will be briefly presented, these are the N-body UNIT simulations (section II.1), the SAGE semi-analytical model (section II.2) and the emission line modelling (section II.3). Next, in section III, we will study a series of figures in which the validity of the galaxy catalogues generated by SAGE will be checked by comparing such properties with observational results. Then in section IV we will examine if the modeling applied to obtain the emission lines of the galaxies is correct. Afterwards in section V the results obtained by studying the two-point correlation function and the bias corresponding to the ELGs in the Euclid range of redshifts will be presented and finally in section VI the conclusions derived from this study will be outlined.

³ <http://www.unitsims.org>

II. THE METHODS

II.1. The UNIT Simulations

As a basis for this work, four gravity only simulations that have been developed within the UNIT project have been employed, the names for the two pairs of simulations that we will use throughout this work is UNITSIM1 (U1) and UNITSIM1-Inverted Phase (U1IP), and UNITSIM2 (U2) and UNITSIM2-Inverted Phase (U2IP). The procedure followed for generating these simulations as well as an analysis of the resulting correlation properties is discussed in [14]. For this particular study we have used the two pairs of simulations in which the code Gadget [50] has been utilized to study the behavior of a total of 4096^3 particles in a volume of $1Gpc^3/h^3$, thus obtaining a mass resolution of $1.2 \cdot 10^9 M_\odot/h$ per simulation particle.

In Chuang et al. [14] besides studying how thanks to the technique developed by Angulo & Pontzen [6] the variance is significantly reduced when studying two point statistics (at least in the range of scales in which baryonic acoustic oscillations occur) it is also explained how halo catalogues have been generated for each of the gravity only simulations using the publicly available ROCKSTAR code [10]⁴.

In this work we have used the halo catalogues resulting from the two pairs of Gadget simulations as a starting point to later use the semi-analytical SAGE code [17] to populate these halos with simulated galaxies (see section II.2) which in turn we will use to study the ELGs (see section II.3) that would ultimately be the type of objects that the Euclid survey would be able to detect.

In the following section we will briefly describe how these semi-analytical models work, focusing in particular on how the SAGE model deals with some of the aspects related to model the galaxy formation and evolution.

II.2. Semi-analytical galaxy modelling via SAGE

Galaxy populations can be included in cosmological simulations using the hydrodynamical equations, however the main advantage of the semi-analytical methods (SAM's) over the hydrodynamic approach is that they are computationally far less intensive, which allows a faster generation of galaxy mock catalogues.

In 1978 D. M. White and M. J. Rees proposed that because DM haloes formed earlier in the history of our universe and then the gas collapsed and cooled down inside those halos forming galaxies [52], an approximate way of dealing with the galaxy evolution problem would be to study how galaxies developed independently but contemplating the possible disturbances caused by the merging between the DM halos they inhabit. From this idea the semi-analytic model technique was born.

Currently many semi-analytic models have been developed, and despite their differences⁵ they all try to implement a series of physical empirically-based laws believed to play a relevant role in galaxy formation using as an underlying canvas the distribution of dark-matter haloes and its merger history. Although now we are just going to briefly describe some of the most relevant physical mechanisms the SAGE code incorporates, the reader can find more detailed information of how this processes are implemented in the SAGE code in the original Croton et al. article [18] or a brief summarized version at the MULTIDARK-GALAXIES paper [36].

II.2.1. Cooling

One of the fundamental phenomena required to explain galaxy formation is gas cooling, which includes several processes whose effects depend on the temperature of the gas. For example in massive halos the gas cools down mainly through bremsstrahlung emission from free electrons and in systems characterized with lower temperatures excitation and de-excitations mechanisms play a more important role; however, it is important to point out that generally all these processes involve the interaction between two particles, hence cooling is generally more efficient

⁴ All the data corresponding to the UNIT project including the halo catalogues is publicly available at <http://www.unitsims.org>

⁵ In the 2010 MULTIDARK-GALAXIES article [36] a through comparative between the semi-analytic codes GALACTICUS, SAG and SAGE is provided. All three models are able to reproduce (but not exactly in the same way) several galaxy properties such as sizes, masses, metallicities, luminosities, etc.

in denser regions where collisions take place more frequently. It is largely because of this process that the baryonic matter segregates from dark matter within a particular halo.

II.2.2. Star formation

As a consequence of the segregation between baryonic and dark matter, the self-gravity of the gas flowing inwards within a halo will start to dominate over the gravity of the halo itself. This causes the gas to collapse further under its own gravity which in return produces higher density concentrations of baryonic matter that cool down even faster. This ends up producing the fragmentation of gas clouds onto high density cores that may eventually lead to the formation of new stars.

However, the precise description in which the stars are formed from the original gas cloud remains unclear, and questions such as: What fraction of the original gas cloud ends up forming stars? Or: What is the mass distribution of the stars which have been generated from a particular gas cloud? have not yet found a satisfactory answer. It is also important to mention that observationally two different modes of star formation have been identified: The quiescent star formation is usually observed in gas disks while starbursts events are believed to be triggered by strong gravitational perturbations in a galaxy which causes the star formation rates to soar.

In SAGE, a Kennicutt-Schmidt type relation [35] is used to calculate the quiescent star formation rate from the cold gas density in the disk. To study the star formation contribution from starbursts, SAGE supposes that this events occur as a consequence of a galaxy-galaxy mergers and estimates the amount of new stars that form from the cold gas the colliding galaxies harbor. It also considers that when the disk in a galaxy becomes unstable (for doing so the Mo et al. model is incorporated [42]) an instability-triggered starburst generates new stars that form in the galaxy bulge and some of the pre-existing stars in the disk are transferred to the bulge in order to make the disc stable again.

II.2.3. Gas ejection and reincorporation mechanisms

The aforementioned cooling processes and the star formation mechanisms suggest that most baryonic material is predicted to end up as cold gas or forming stars, however observations show that only a relatively small fraction of baryonic matter finishes forming up stars or in cold gas clouds, this is the reason why in galactic evolution models certain processes must be considered to mitigate these effects.

There are mainly two mechanism which may produce the reheating of the gas and its ejection from a given galaxy:

- ☞ Supernovae explosions are believed to produce enough energy to reheat and eject a notorious fraction of gas in a galaxy in what is known as supernovae galactic winds. In SAGE this process is dealt with as follows: At first cold gas is blown out of the disc and into the hot halo, then the thermal energy of the supernova explosion is added to the energy of the hot halo, and if the total exceeds a particular binding energy of the hot halo, part of the gas in it will be ejected and will become part of an outside gas reservoir. The gas in this outside reservoir will slowly get reincorporated back in the hot halo of the galaxy at a specific rate.
- ☞ It has been determined observationally that almost all galaxies contain a supermassive black hole (SMBH) in their center. These supermassive black holes are believed to experience active accretion phases during their lives in which vast amounts of energy are depleted which, at the same time, induces galactic winds where gas clouds are reheated and even ejected. During these supermassive black hole accretion phases the galaxy is said to have a active galactic nuclei (AGN). The way in which this phenomena is treated within SAGE is explained in Croton et al. [18] and is important to point out that the the mass of the SMBH in the galaxies (which is a crucial parameter in these kind of processes) is affected by the galaxy mergers and the disk instabilities.

II.2.4. Chemical Evolution

It is believed that the baryonic matter in the universe was mainly hydrogen ($\sim 75\%$) and helium ($\sim 25\%$) until the first stars formed, when nuclear reactions inside those stars caused the development of heavier elements. After a star dies, the elements of which it is composed of are scattered across the galaxy through galactic winds, enriching the interstellar medium with elements heavier than hydrogen and helium that may end up forming new stars.

In SAGE, the metal treatment is based upon the De Lucia et al. article [24] and considers the so-called instantaneous recycling approximation where a fraction of the mass of newly formed stars is immediately recycled back to the cold gas disk assuming that a certain fraction of the mass conforming the new star has been converted into metals [18].

II.2.5. Merger treatment

As we have already mentioned, one of the most important processes to describe galaxy evolution is the merging history of the dark matter halos they inhabit. This is due to the fact that these events make possible the merging between galaxies inside the halos that come together, which obviously alters greatly their characteristics when the sizes of the merging galaxies are more or less similar; if one of the galaxies that merges is considerably bigger than the other, the contents of the smaller one will be incorporated into the bigger one, whose properties will not change dramatically. Also the merging between dark matter halos can provoke starburst events due to the gravitational perturbations they cause.

For any SAM to reproduce the evolution of a set of galaxies a “merger tree” is required, this object describes the merging history of dark matter halos and can be obtained tracing the positions of dark matter halos in a gravity only simulation. In SAGE once the occurrence of a galaxy–galaxy merger has been identified, the merger is treated as “major” when the ratio between baryonic mass of the bigger galaxy to the baryonic mass of the smaller galaxy is greater than 0.3, in this case the disks of both galaxies are destroyed and all stars are combined to form a new spheroidal galaxy; when the 0.3 threshold is not reached, the merger will be treated as “minor” and the satellite stars will be added to the central galaxy bulge [18].

II.2.6. Calibration method

It is imperative to point out that SAGE (as any other SAM) needs to determine some of the parameters it uses from a certain set of external samples. For doing so, it primarily calibrates several parameters using the fitting data over the $z = 0$ stellar mass function and secondarily the fitting of other relevant relations such as the stellar metallicity–mass relation, the baryonic Tully-Fisher relation, the black hole–bulge mass relation and the cosmic star formation rate density. So it’s expected that certain relations such as the stellar mass function or the stellar metallicity–mass relation are well reproduced, at least at $z = 0$.

II.3. Emission-line galaxy modelling

Once we have populated the dark matter halos from the UNIT simulations with the semi-analytical galaxies generated by SAGE we are going to obtain values for the intensity of some of the most relevant emission lines such as H_α , [OIII]4959, [OIII]5007, [NII]6548 and [NII]6584 corresponding to each one of the simulated galaxies⁶. However, in this particular study we will only focus on H_α .

🌀 GET_EMLINES code:

In order to reproduce the intensity of H_α emission lines of our galaxies we have used the GET_EMLINES code (Orsi et al. [44])⁷. This code is based on the algorithm MAPPINGS-III described in Groves et al. [31] and Allen et al. [2] which relates the ionisation parameter of gas in galaxies, q , to their cold gas metallicity Z_{cold} as:

$$q(Z) = q_0 \left(\frac{Z_{\text{cold}}}{Z_0} \right)^{-\gamma}, \quad (2)$$

where q_0 is the ionisation parameter of a galaxy that has cold gas metallicity Z_0 and γ is the exponent of the power law. We adopted the suggested values of $q_0 = 2.8 \times 10^7 \text{ cm s}^{-1}$ and $\gamma = 1.3$ which were found to yield H_α

⁶ It is important to point that semi-analytical models do not provide directly emission-line properties, this is the reason why we need to resort to this kind of modeling.

⁷ the GET_EMLINES code is the publicly available at: https://github.com/aaorsi/get_emlines

luminosities for star-forming galaxies in good agreement with observations [44]. In order to obtain the value for the the cold gas metallicity (Z_{cold}), we compute the ratio between the cold gas mass in metals and the cold gas mass [53] considering both the bulge and disk components:

$$Z_{\text{cold}} = \frac{M_{Z_{\text{cold}}}}{M_{\text{cold}}}. \quad (3)$$

The other relevant parameter required to compute the intensity of the emission lines with this particular code is the total star formation rate (again we will use the sum of the values for the disk and the bulge). Note that SAGE provides this quantity averaged over the last time-step but the model ideally requires as inputs the instantaneous SFR and cold gas metallicity of galaxies, nevertheless, Favole et al. [30] have shown that the differences are negligible.

To be able to properly compare our results to observations we convert the luminosities to fluxes and also apply a dust extinction model to the luminosities obtained for our simulated galaxies.

☞ Dust attenuation:

The particular dust model that we have used is described in great detail in Favole et al. [30], but we also summarize it here. The attenuation from interstellar dust is added to the intrinsic H α luminosity using the following expression:

$$L(\lambda_j)^{\text{att}} = L(\lambda_j)^{\text{intr}} 10^{-0.4A_\lambda(\tau_\lambda^z, \theta)}, \quad (4)$$

where the attenuation coefficient, as a function of the galaxy optical depth τ_λ^z and the dust scattering angle θ , is defined as [25, 30, 34, 45]:

$$A_\lambda(\tau_\lambda^z, \theta) = -2.5 \log_{10} \frac{1 - \exp(-a_\lambda \sec \theta)}{a_\lambda \sec \theta}; \quad a_\lambda = \sqrt{1 - \omega_\lambda} \tau_\lambda^z \quad (5)$$

Where ω_λ is the dust albedo. We assume $\cos \theta = 0.30$ and $\omega_\lambda = 0.56$, meaning that the scattering is not isotropic but forward-oriented, and about 60% of the extinction is caused by scattering.

The galaxy optical depth is defined in terms of the cold gas metallicity Z_{cold} (see Eq.3) as [23, 33]:

$$\tau_\lambda^z = \left(\frac{A_\lambda}{A_V} \right)_{Z_\odot} \left(\frac{Z_{\text{cold}}}{Z_\odot} \right)^{1.6} \left(\frac{\langle N_H \rangle}{2.1 \times 10^{21} \text{ atoms cm}^{-2}} \right), \quad (6)$$

And we will assume that $Z_\odot = 0.0134$ which is the value obtained for the Milky Way [7]. We are also going to assume that the extinction law can be described as in Cardelli et al. [12]:

$$\left(\frac{A_\lambda}{A_V} \right) = a(x) + b(x)/R_V, \quad (7)$$

Where $x \equiv \lambda^{-1}$, $R_V \equiv A_V/E(B-V) = 3.1$ is the ratio of total to selective extinction for the diffuse interstellar medium in the Milky Way, and the coefficients $a(x)$ and $b(x)$ are given by:

$$\begin{aligned} a(x) &= 1 + 0.17699 y - 0.50447 y^2 - 0.02427 y^3 + \\ &\quad 0.72085 y^4 + 0.01979 y^5 - 0.77530 y^6 + 0.32999 y^7, \\ b(x) &= 1.41338 y + 2.28305 y^2 + 1.07233 y^3 - 5.38434 y^4 \\ &\quad - 0.62251 y^5 + 5.30260 y^6 - 2.09002 y^7, \end{aligned} \quad (8)$$

with $y = x - 1.82$. Also, the quantity $\langle N_H \rangle$ in Eq.6 is the mean hydrogen column density defined as [23, 33]:

$$\langle N_H \rangle = \frac{M_{\text{cold}}^{\text{disc}}}{1.4 m_p \pi (a R_{1/2}^{\text{disc}})^2} \text{ atoms cm}^{-2}, \quad (9)$$

where $M_{\text{cold}}^{\text{disc}}$ is the cold gas mass of the disc, $m_p = 1.67 \times 10^{-27}$ kg is the proton mass, $a = 1.68$ is a normalisation factor that makes the column density representing the mass-weighted mean column density of the disc, and $R_{1/2}^{\text{disc}}$ is the half-mass radius of the disc.

III. THE SAGE GALAXIES

The aim of this section is to compare the results that correspond to the semi-analytical galaxies obtained by applying the SAGE code to the halo catalogues from the UNIT simulations with a series of available observational data.

Since there is a greater amount of observational data for the properties of galaxies for redshifts close to zero, in most of the figures we present we will study whether the results are valid in this low redshift regime. Whenever possible we use observational data at higher redshifts that will allow us to verify if our results reasonably comply with the properties observed for galaxies at a similar range of redshifts to those that the Euclid mission will cover.

Our prime focus here lies on properties that will eventually be used to calculate emission lines, that is stellar mass, star formation rates, and metallicities. For additional validation plots we refer the reader to App.A.

III.1. Stellar Mass Function

The stellar mass function (SMF) is one of the most significant properties that can be inferred from galaxy surveys since this function represents the number of galaxies normalized to the volume of the survey/simulation and to the width of the bins employed as a function of the mass that is found in those particular galaxies forming stars. This is why the SMF is often employed for calibrating semi-analytical models such as SAGE.

In the main panel of Fig. 2 the results obtained for the SMF computed from the SAGE galaxies modelled over the UNITSIM 1 simulation [13] is presented for three different redshifts $z \in (0.0, 1.710, 2.695)$. Together with the results obtained from our simulation, a series of observational results obtained for a range of redshifts similar to those simulated are also represented in the same figure: the compilation for redshift $z = 0$ is taken from the so-called ‘CARNage calibration’ data set described in great detail in Sec. 3.3 and the Appendix A of Knebe et al. [37], the observations for the higher redshifts are taken from Davidzon et al. [20] and are based on the UltraVISTA near-infrared survey of the COSMOS field.

In the bottom panel of Fig. 2 the variation between the different simulations is derived from the differences between the results obtained for the SMF computed from the data coming from UNITSIM1 (U1) and the SMF obtained for the other three simulations carried out in the UNIT project (U1 inverted phase, U2 and U2 inverted phase) are represented according to the following formula:

$$\delta(U_i, U_j) = \frac{\text{SMF}(U_i)}{\text{SMF}(U_j)} - 1 \quad (10)$$

For all the simulations conducted the results produced for the SMF are compatible with the observational data collected at different redshifts. This outcome is in line with previous results such as those obtained in Favole et al. [30] and Asquith et al. [8]. However, it is important to notice that whereas the results obtained at redshift $z = 0$ agree seamlessly with the observational data, when studying the behavior at higher redshifts, certain discrepancies start to show up. For stellar masses below $10^{11} M_{\odot}$ approximately the SMF calculated with the UNIT-SAGE galaxies overestimates the observational value recorded for the number of galaxies with a specific mass, while the opposite occurs for masses higher than $10^{11} M_{\odot}$.

Despite the discrepancies we have mentioned, the results obtained are reasonably accurate for us to say that the modelled galaxies fairly depict the behaviour of the SMF that could be expected in the redshift range for which Euclid is expected to operate. Additionally, the discrepancies observed here for high redshifts are similarly found when studying the SMF produced by other semi-analytical models different from SAGE as discussed in Asquith et al. [8].

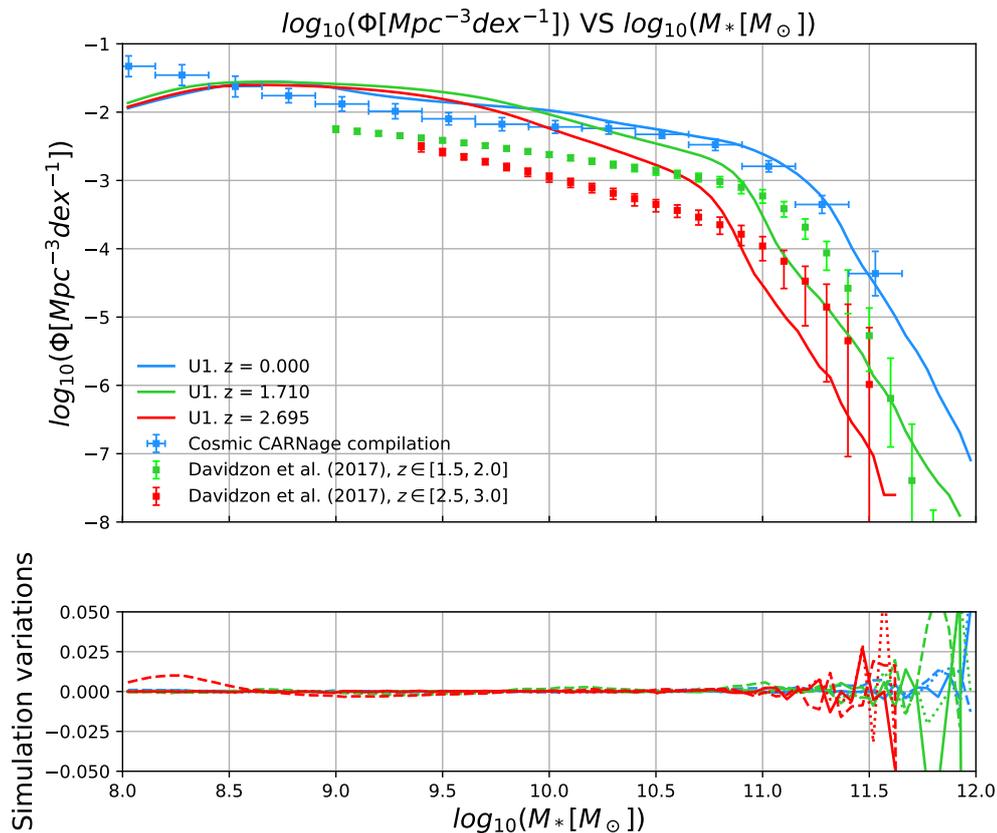


Figure 2: Stellar mass function.

Another important aspect worth mentioning in this section is that due to resolution limitations in our simulations, galaxies whose stellar mass is lower than $10^9 M_\odot$ should not be considered. Therefore to produce the results presented in the following sections we will discard all those galaxies whose mass is inferior to this threshold.

III.2. Star Formation

III.2.1. The Star Formation Rate Function

Another fundamental property that can be studied when analyzing galaxy statistics is the star formation rate function (SFRF). This feature provides a way to study the number of galaxies (normalized by volume of the simulation/survey and by the width of the bins employed) that have a given star formation rate.

In the main panel of the Fig.3 the SFRF calculated from the SAGE-UNIT galaxies at redshift $z = 0.142$ is displayed. Next to this result, we have also plotted the observational values for the SFRF calculated by Gruppioni et al. [32] for galaxies with $z \in (0, 0.3)$. In the bottom panel of Fig.3, the variations between the SFRF obtained for U1 at redshift $z = 0.142$ (shown in the main panel) and those generated from the SAGE galaxies for the other three UNIT simulations (U1 inverted phase, U2 and U2 inverted phase) have been represented in a completely analogous fashion as in the bottom panel of Fig.2.

In Fig.3 it can be seen how the results obtained for the simulated SFRF are compatible with the observational data displayed for a similar redshift range. The results presented here are also consistent with those presented for the SFRF by Favole et al. [30], but it is worth mentioning that for SFR values greater than $10^{-6} M_\odot/\text{yr}$ approximately, the number of galaxies generated with SAGE seems to underestimate the number of galaxies observed in that particular SFR range.

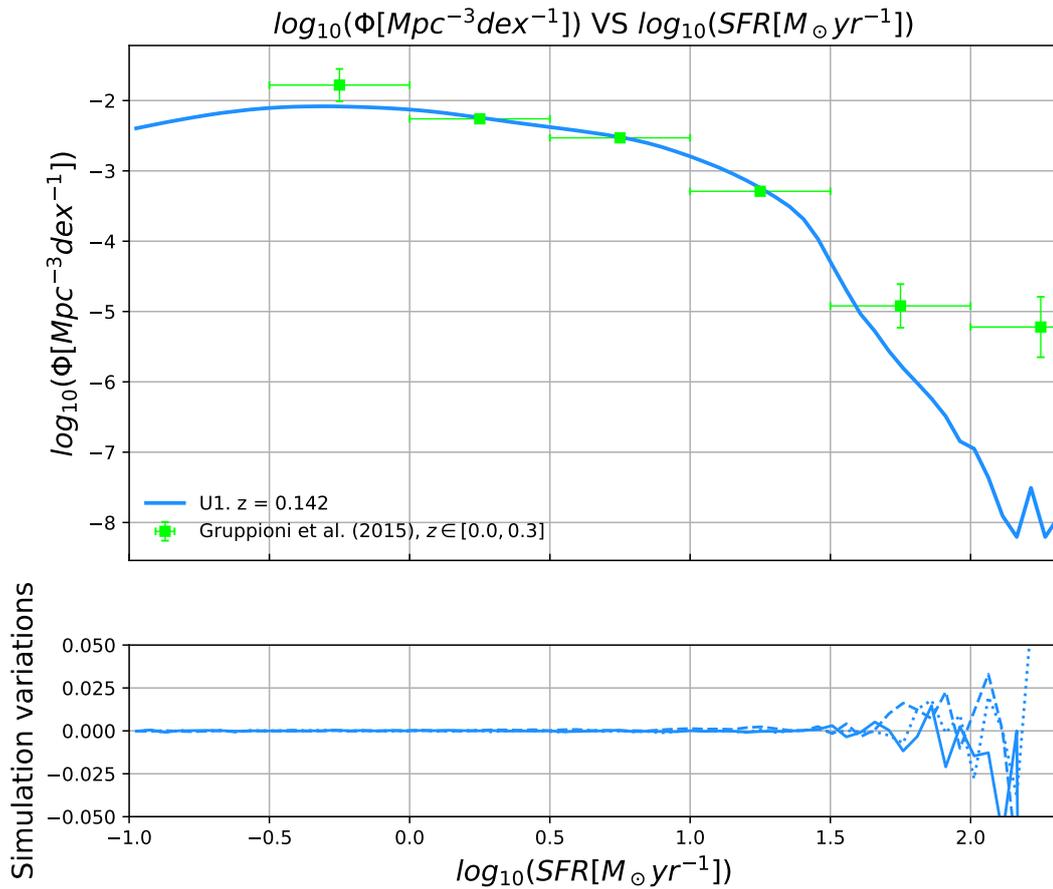


Figure 3: Star formation rate function.

Finally it is worth noting that we have studied how the SFRF evolves with redshift and we have been able to observe that the maximum for this function moves towards higher SFR values progressively, reaching for example the value for $\text{SFR} \approx 3M_{\odot}/\text{yr}$ at a redshift $z \approx 2$. This behavior is within expectations, and we expect that our simulated galaxies are also correctly emulating the behavior for the SFR of the real galaxies that Euclid will be able to observe within the $0.9 < z < 1.8$ range.

III.2.2. The Specific Star Formation Rate to Stellar Mass Relation

The specific SFR (sSFR) gives us information about the number of stars that form per unit of stellar mass for a given galaxy. This property is also commonly examined in many studies which attempt to estimate its value for galaxies at low redshifts.

In the main panel of Fig.4 the sSFR calculated from the UNITSIM1-SAGE galaxies is represented. To calculate this value, the median of the values obtained for sSFR within a series of bins along the x-axis has been taken (the fundamental plot from which this figure is constructed is a scatter in which for each galaxy a point is plotted with y-coordinate corresponding to the sSFR of that galaxy and x-coordinate corresponding to its SMF). In addition to the curve corresponding to the median value of the sSFR as a function of the star mass, on the main panel of Fig.4 the reader can appreciate a contour plot which helps to visualize the distribution of galaxies in this plane.

As customary, in the bottom panel of the Fig.4 the variations between simulations with respect to the other SAGE-UNIT simulated galaxies has been represented.

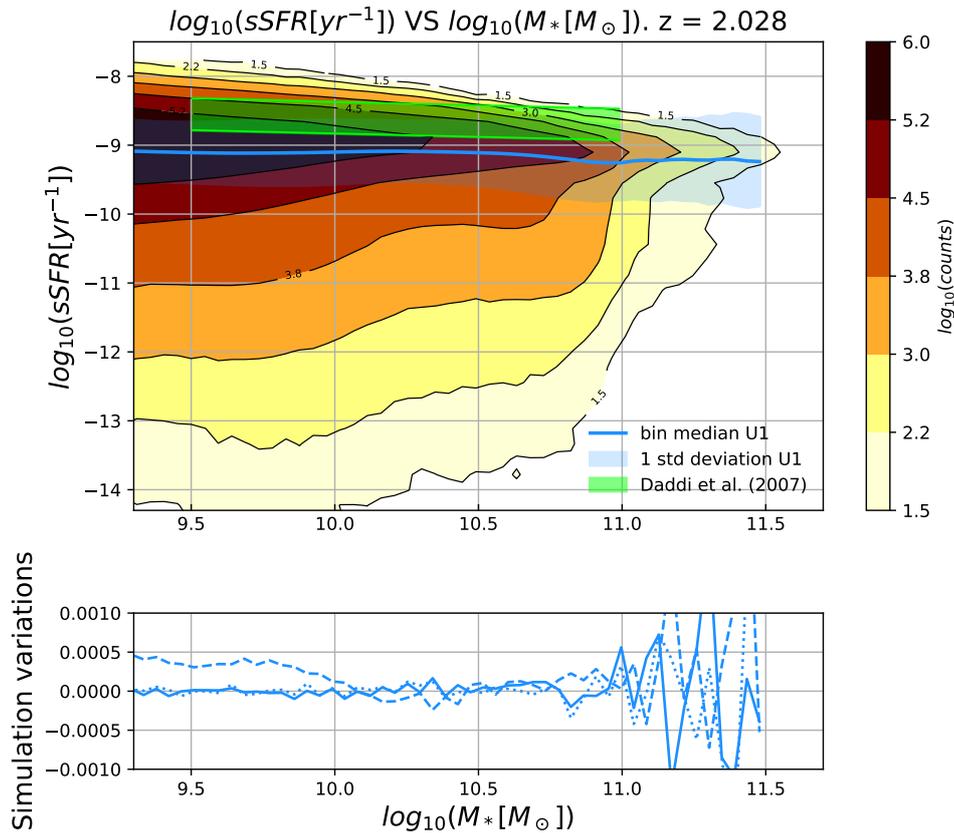


Figure 4: Specific star formation rate against stellar mass.

As we can see in the main panel of Fig.4, we compare the results obtained for the simulated sSFR with observational data extracted from Daddi et al. [19]. The observational data for the sSFR is in good agreement with the results found with our simulation at least for redshift $z \approx 2^8$.

Again, based on the results obtained, we can say that the galaxies that have been simulated also seem to accurately reproduce the behaviour of the sSFR that would be expected for a sample of real galaxies in Euclid’s operating range of redshifts.

III.3. The Mass-Metallicity Relation

Another aspect that we cannot ignore when studying the evolution of galaxies is their chemical composition, since depending on the fraction of metals that a galaxy may contain, its SFR may be substantially modified due to the fact that a higher metal content favours cooling mechanisms. In addition, this property is particularly crucial for our work because in order to calculate the intensity of the H_α emission line using the program described in Orsi et al. [44] it is necessary to provide the code as input the cold gas mass of the galaxies.

Since star formation is regulated by the collapse of cold gas clouds, in Fig.5 we study the relation that exists between the total mass of metals contained in such clouds and the total mass of cold gas in a given galaxy throughout the parameter Z_{cold} which is defined in Favole et al. [30] as:

$$Z_{\text{cold}} = 8.69 + \log_{10} \left(\frac{M_{Z,\text{cold}}}{M_{\text{cold}}} \right) - \log_{10} 0.0134 \quad (11)$$

⁸ Although not represented here, we have also verified that the results obtained from the simulated galaxies are in full agreement with observational values for the sSFR as a function of stellar mass found for redshift $z = 0$ in Elbaz et al. [28]. These results, in turn, are also compatible to those shown in Favole et al. [30] for redshift $z = 0.1$

In the main panel of Fig.5 the curve corresponding to the median taken for different bins over the two-dimensional distribution of points is represented as in previous sections and in the bottom panel the variations with respect to the rest of the simulations can be appreciated.

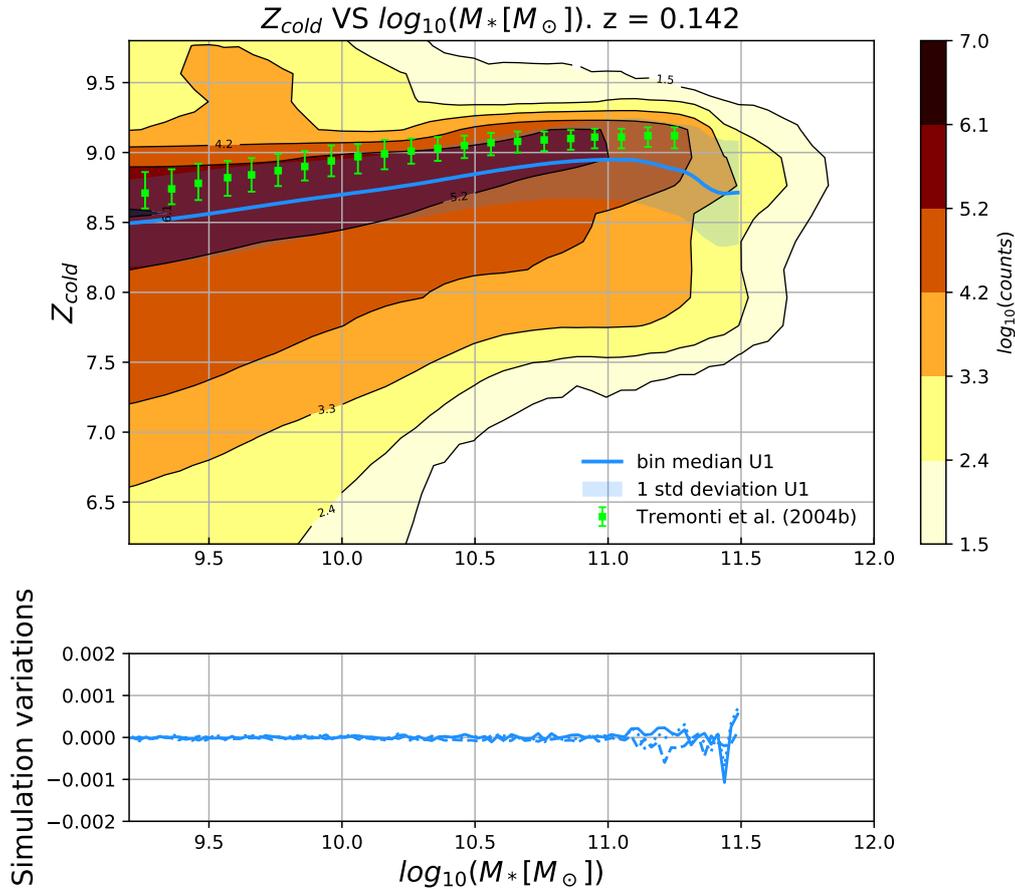


Figure 5: Cold gas mass in metals against stellar mass.

Along with the data coming from the simulations at redshift $z = 0.142$ we have also represented the observational data presented in Tremonti et al. [51] at redshift $z \approx 0.1$. Considering the compatibility between the observational results and the simulated ones, we can assert that in the simulated galaxies we are also managing to correctly reproduce the amount of metals that can be found as part of the cold gas.

IV. SAGE'S EMISSION-LINE GALAXIES (ELGS)

IV.1. Abundance of dust-attenuated ELGs

The model presented in II.3 provides us with the relevant luminosities in H_α for all our SAGE galaxies. However, to properly compare them to observations at a particular redshift they were converted to fluxes as well as being passed through an appropriate dust attenuation model (see II.3).

To verify the credibility of our (attenuated) ELGs we study in Fig.6 the cumulative distribution functions of the number of ELGs per deg^2 as a function of H_α flux. The left panel shows the original fluxes for a series of redshifts on the interval $z \in [1.0, 2.0]$ whereas the right panel shows the same distributions after applying the dust correction.

We decided to only show the results for UNITSIM1 as the other three ELG catalogues give indistinguishable results. The data points in both panels are observations taken from Colbert et al. [15]. While it is obvious how the

dust model shifts the contribution to the cumulative number counts towards lower fluxes (and into better agreement with observations), we also find that our modelled ELGs lack objects at high fluxes. In any case, we observe that with decreasing redshift the number of ELGs for a given flux threshold increases.

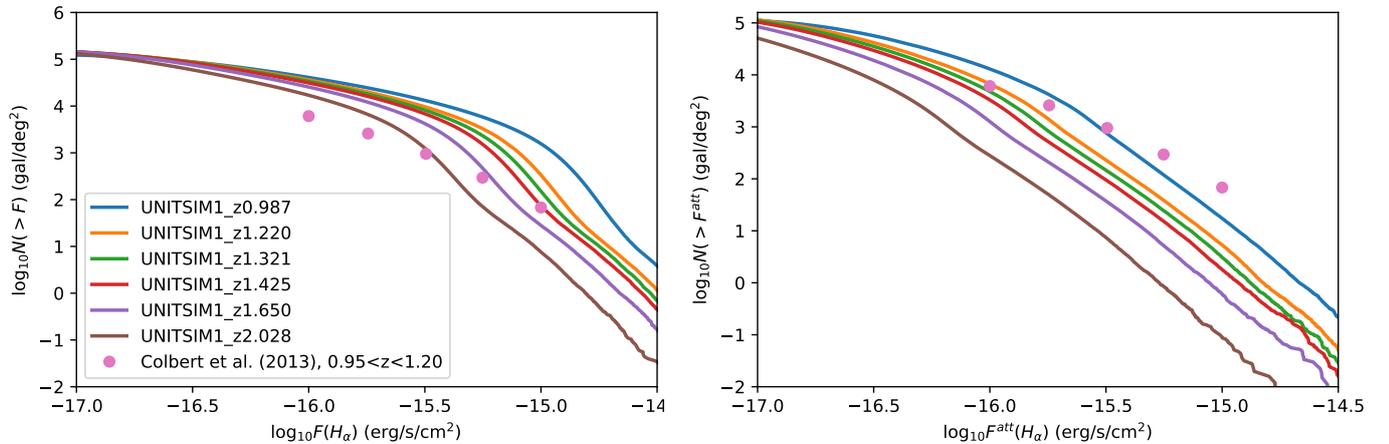


Figure 6: Cumulative counts of galaxies per deg^2 as a function of H_α flux in comparison to the observational data of Colbert et al. [15]. The left panel shows un-attenuated fluxes while in the right panel we are representing the same figure but considering the dust attenuated counts as described in Sec. II.3. Only results for UNITSIM1 are shown for clarity since the results obtained for the other three simulations that we have generated exhibit essentially the same behaviour.

IV.2. Abundance evolution of flux-selected ELGs

Throughout the remainder of the paper we will use attenuated fluxes and apply lower limits to them. Our reference flux threshold is $F_{\text{cut}}^{\text{ref}} = 2 \times 10^{-16} \text{ erg/s/cm}^2$ across all redshifts which corresponds to the limit of the Euclid satellite; we refer to this model as ‘RefMod’.

At this point it is necessary to mention that there are currently several works that make forecasts for the number of objects (galaxies) that the Euclid mission will be able to observe as a function of redshift. One of the most relevant works and the one that is usually employed currently as a reference to predict the number density of galaxies by deg^2 that the Euclid mission will be able to detect for different redshifts is the one presented by Pozzetti et al. [46].

We show in Fig. 7 the redshift evolution of the number density of ELGs for this particular model in comparison to observational data of Colbert et al. [15]. We also show two of the three models of Pozzetti et al. [46] which were constructed by fitting to observed luminosity functions from H_α surveys: we show model #1 and #3 for a flux cut of $2 \times 10^{-16} \text{ erg/s/cm}^2$ ⁹. These two models are considered the most optimistic (model #1) and most pessimistic (model #3) for Euclid and we refer to them as ‘PozMod1’ and ‘PozMod3’, respectively. We can see how our ‘RefMod’ ELGs follow the same trends as the Pozzetti et al. [46] models, but show a substantial lack of objects at higher redshifts.

This discrepancy is also reflected in Tab. I where we list as a function of redshift the number density of ELGs in our reference model ‘RefMod’ (as averaged over the four UNIT simulations, also providing the standard deviation). In order to correct for this deficit we are going to adjust our flux thresholds at each redshift so that we are matching the number densities of Pozzetti et al. [46] also provided in that Table in the columns with the headings ‘PozMod1’ and ‘PozMod3’ respectively: for a fixed redshift, we vary the flux threshold F_{cut} when selecting our UNITSIM ELGs until we match that particular number density. Note that the Pozzetti et al. [46] predictions are for the Euclid reference

⁹ The Pozzetti et al. [46] data can be downloaded from here: <http://www.bo.astro.it/~pozzetti/Halpha/Halpha.html>

flux cut of $F_{\text{cut}} = 2 \times 10^{-16}$ erg/s/cm². This is done for all four UNITSIM catalogues and the means are listed in the table; we omit error estimates as they are below the reported accuracy.

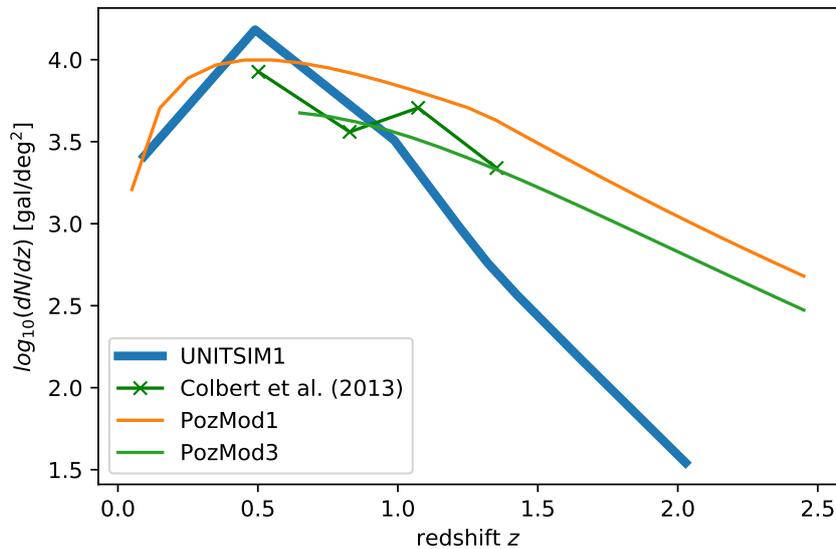


Figure 7: Redshift evolution of dN/dz (measured as N/deg^2) for the RefMod (i.e. universal flux cut at 2×10^{-16} erg/s/cm²) in comparison to the observational data of Colbert et al. [15] and model #1 (PozMod1) and #3 (PozMod3) of Pozzetti et al. [46] for the same flux threshold. Only UNITSIM1 ELGs are shown for clarity since the results obtained for the other three simulations that we have generated exhibit essentially the same behaviour.

z	RefMod ($F_{\text{cut}} = 2$)		PozMod1		PozMod3	
	$\langle dN/dz \rangle$	σ	$\langle dN/dz \rangle$	F_{cut}	$\langle dN/dz \rangle$	F_{cut}
0.490	15262	30	9946	2.85	–	–
0.987	3238	12	7353	1.37	3779	1.88
1.220	957	7	5097	1.13	2518	1.50
1.321	577	4	4281	1.05	2148	1.35
1.425	370	3	3447	0.98	1817	1.22
1.650	153	3	2253	0.84	1279	1.00
2.028	35	1	1006	0.67	616	0.77

Table I: The two columns for the reference model ‘RefMod’ are listing the average number density dN/dz of ELGs (measured as N/deg^2) and its standard deviation when applying a redshift-independent flux cut of 2×10^{-16} erg/s/cm² to our dust-attenuated ELGs. The two columns for each of the Pozzetti et al. [46] models ‘PozMod1’ and ‘PozMod3’, respectively, give the predicted number density of ELGs (taken from their Table 3) and the flux threshold required to match that number for our SAGE ELGs, again as an average over all four UNIT simulations (the standard deviation is smaller than the reported accuracy).

We like to close this section with the remark that the most recent study of the number density evolution of H_α ELGs indicates a possible decline beyond redshift $z \sim 1.4$ Bagley et al. [9] though not as pronounced as the dip found for our SAGE ELGs. In any case, this results seem to indicate that even model #3 of Pozzetti et al. [46] might overpredict high- z ELGs densities. Furthermore, it is important to point out that in Pozzetti et al. [46], predictions for the galaxy number density as a function of redshift for the reference flux cut of Euclid are examined using other semi-analytical codes and all seem to produce similar results to those obtained with our modelling.

V. CORRELATION PROPERTIES OF ELGS

We will now change the focus to the clustering properties of our SAGE ELGs. To this extent we present the results derived from studying the two-point correlation function (2pcf) using the positions of the modeled ELGs as well as 10^7 randomly selected dark matter particles from the total 4096^3 particles present in each of the UNIT dark matter only simulations¹⁰. From the results obtained for the 2pcf of the ELGs and for the catalogues of collisionless particles of the UNIT simulations we can infer the bias between both distributions and study its evolution with redshift.

To produce these results we will use the CUTE software presented in Alonso [3] and take into account that for each of the redshifts of interest we apply different flux cuts to make the ELGs selection corresponding to the different predictions for what Euclid would be able to observe according to the values presented in Tab.I. In addition, for the results that we will present throughout this section, we have taken the average of the values computed for the 2pcf over the four simulations at our disposal.

In the top left panel of Fig.8 we present the average over the four UNIT simulations for the 2pcf computed for the RefMod ELGs (dashed lines) and dark matter (solid lines). The lower left panel of the same figure shows the bias $b(r)$ defined as [29, 43]:

$$b(r) = \sqrt{\frac{\xi_{\text{ELGs}}(r)}{\xi_{\text{DM}}(r)}}, \quad (12)$$

The bias is further averaged over all the values of r excluding the value corresponding to the smallest distance considered here since, as we will see in Fig.9, for such a small distances the bias exhibits a non-linear behavior. This averaged bias is presented as dashed straight lines. It is obvious that the data for this particular model becomes rather noisy at high redshifts due to the very low number of objects above the reference flux cut of $F_{\text{cut}} = 2 \times 10^{-16}$ erg/s/cm².

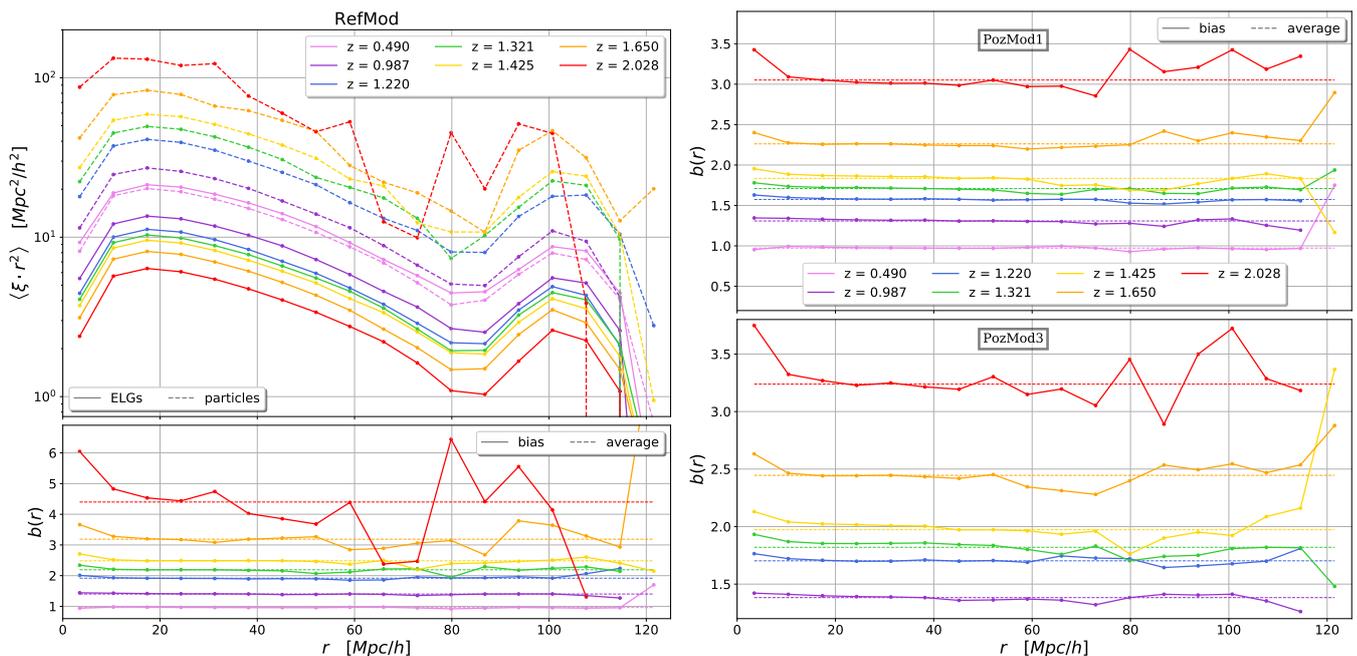


Figure 8: Top left panel: 2pcf of SAGE ELGs with flux greater than 2×10^{-16} erg/s/cm² (RefMod, dashed lines) and collisionless trace particles (solid lines) for various redshifts. Bottom left panel: associated bias to the RefMod ELGs as defined by Eq.12. Top right panel: bias for the PozMod1 ELGs. Bottom right panel: bias for the PozMod3 ELGs.

¹⁰ The reason why we make this sampling is because calculating the two-point correlation function is a very expensive computational task and selecting this number of particles randomly for the dark matter only simulations allows us to calculate the 2pcf with sufficient precision so that there are no substantial differences when comparing the result obtained with the 2pcf produced if the total number of DM particles would have been employed.

An equivalent analysis to the one presented above has been conducted for the 2pcf obtained from the selected ELGs by applying the flux cuts that match model #1 (PozMod1) and #3 (PozMod3) of Pozzetti et al. [46] number densities (see Tab.I). The results derived for the bias of the ELGs are presented for both models and for the different redshifts of interest in the two right panels of Fig.8. Once again we can see how due to the reduced number of galaxies for both models at redshift 2.028 we get fairly noisy results. In Fig.8 we have presented the bias and its averaged value (excluding the value corresponding to the smallest distance) over all scales, here we can observe how the bias is consistently scale invariant for scales higher than $r > 20 h^{-1}Mpc$, below this threshold value we expect that the mixture contribution between the one- and two-halo terms will introduce non-linear effects which in turn will cause the bias to no longer behave independently with the scale.

In order to study how the bias varies as a function of the distance we have decided to examine the behaviour of the bias at small scales ($0.1 < r < 20 h^{-1}Mpc$) for the three presented models and for the different redshifts of interest with greater detail. The results in this case are presented in Fig.9 in an analogous fashion to the results shown in the right panels of Fig.8. We observe that for redshifts $z < 2$ the bias remains again constant down to scales $r \approx 3 h^{-1}Mpc$ and then start to mildly drop, therefore it is actually around this value that we expect the contribution from the 1-halo term to start to become relevant. However, this behaviour weakens for large redshifts and even reverses for $z = 2$.

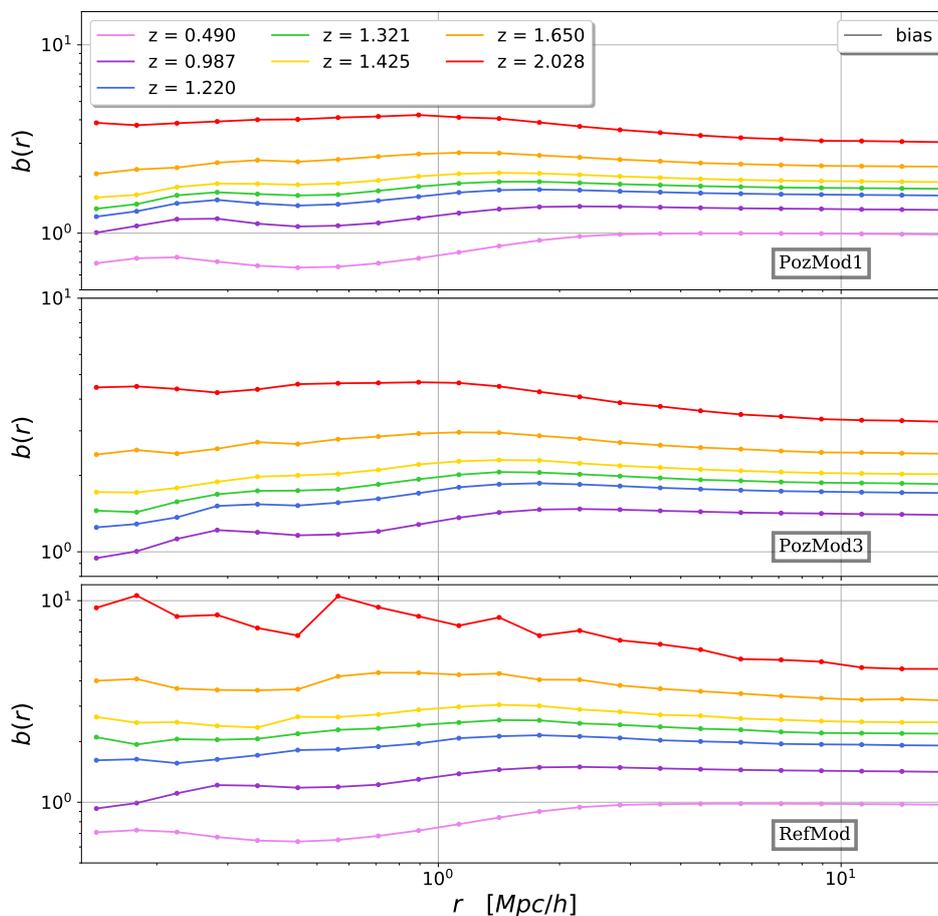


Figure 9: The bias $b(r)$ focusing on small scales ($r < 20 h^{-1}Mpc$). The upper panel shows the bias computed for PozMod1 ELGs, the middle one for PozMod3, and the lower panel for RefMod.

To conclude this section we present in Fig.10 the evolution of the bias averaged for large scales (obtained from the results presented in Fig.8) as a function of redshift for the three models of interest. This figure is accompanied by Tab.II that lists exactly the numbers used for it. We find that for all our models – even for the ones where we matched the number densities to the Pozzetti et al. [46] predictions – the bias systematically increases with redshift which is in agreement with the findings of Favole et al. [29], which reflects the fact that the number of galaxies with high H_α

z	RefMod	PozMod1	PozMod3
0.490	0.96	0.97	-
0.987	1.4	1.31	1.38
1.220	1.92	1.57	1.7
1.321	2.19	1.71	1.82
1.425	2.48	1.83	1.97
1.650	3.19	2.26	2.45
2.028	4.4	3.05	3.24

Table II: Bias values averaged for scales larger than $5 h^{-1} Mpc$ computed for the three models of interest (RefMod, PozMod1, PozMod3) corresponding to the dashed lines presented in the three bias panels of Fig.8 and for the different redshifts studied.

fluxes decreases as we observe higher redshifts and therefore it is expected that most of the galaxies detected by the Euclid mission would be in the lower range of redshifts for which it is designed to operate. However, in Favole et al. [29] work the bias increases more mildly since the SDSS redshift range studied there is very much reduced compared to ours. Further, for SDSS they only studied the large-scale linear bias on scales larger than $8 h^{-1} Mpc$ and not the non-linear regime probed here, too.

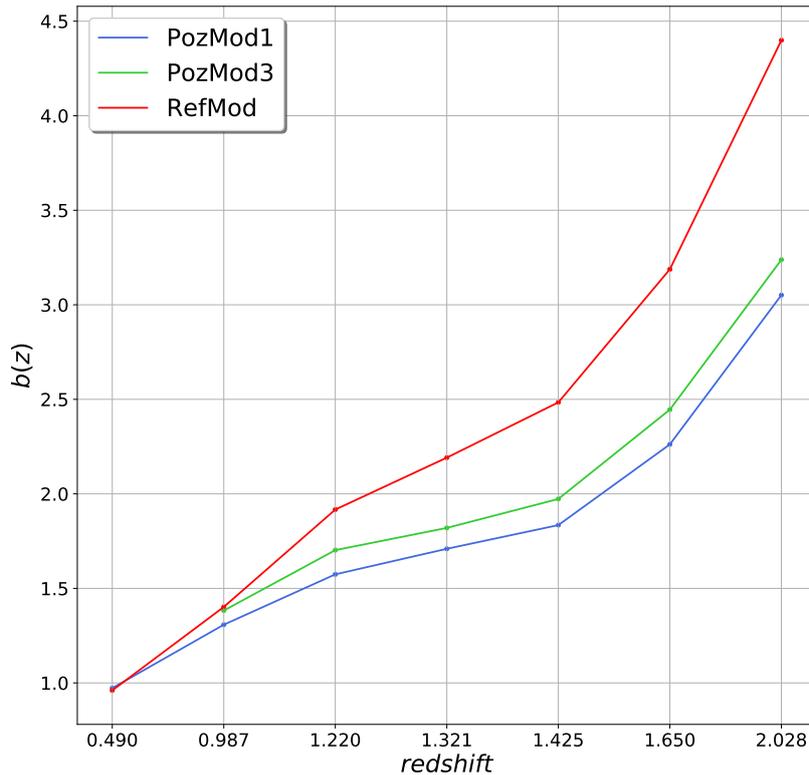


Figure 10: Bias values averaged for scales larger than $5 h^{-1} Mpc$ computed for the three models of interest (RefMod - red, PozMod1 - blue, PozMod3 - green) corresponding to the dashed lines presented in three bias panels of Fig.8 as a function of the studied redshifts.

VI. CONCLUSIONS

In this work we explained how we have generated four synthetic catalogues of H_α ELGs by modelling the emission lines of galaxies (using the GET_EMLINES code with an additional dust attenuation model) and employing the properties calculated for the semi-analytical galaxies generated from the SAGE model over the UNIT simulations. As argued throughout Sec.III and App.A of this paper, the properties associated with our UNIT-SAGE galaxies appear to reproduce as best as one can expect the behavior observed for real galaxies in the range of redshifts $0 \leq z < 2$. Because of this, we will be able to make predictions for the galaxy statistics that the Euclid experiment is expected to obtain for redshifts between $0.9 < z < 1.8$.

As discussed in Sec.II.3 of this work, the flux computed for the H_α lines produced taking into account the dust attenuation contribution seems to adequately reproduce the corresponding observational results. However, when studying the density of galaxies per deg^2 with fluxes greater than 2×10^{-16} erg/s/cm² as a function of redshift we observe that the density obtained for the UNIT-SAGE ELGs might underestimate the expected density of galaxies observed from redshift $z \sim 1.4$ onwards, although the most recent study by Bagley et al. [9] suggests that the observational value for this quantity should be closer to our results and therefore that the densities predicted by Pozzetti et al. [46] can give overestimated galaxy densities at high redshifts.

When studying the bias for the selected ELGs by applying the flux cuts for the three different models considered listed in Tab.I, an scale independent bias has been obtained for distances $r < 3 h^{-1}Mpc$. However, below this threshold value we have been able to observe by doing an analysis exclusively for distances below $20 h^{-1}Mpc$ that non-linear related effects begin to arise. The values obtained for the bias averaged at scales free of non-linear effects ($r < 5 h^{-1}Mpc$) are listed in Tab.II.

These predictions allow, on a qualitative basis, to impose an approximate scale threshold ($r \sim 3 h^{-1}Mpc$) around which the formation of structures can be assumed to transition from a linear regime to a non-linear regime, and, on a quantitative basis, the values provided for the galaxy bias can be useful in future works trying to predict the distribution properties of galaxies with H_α emission within the range of redshifts studied here.

-
- [1] Abbott T. M. C., et al., 2018, *The Astrophysical Journal Supplement Series*, 239, 18
 - [2] Allen M. G., Groves B. A., Dopita M. A., Sutherland R. S., Kewley L. J., 2008, *The Astrophysical Journal Supplement Series*, 178, 20–55
 - [3] Alonso D., 2012, CUTE solutions for two-point correlation functions from large cosmological datasets ([arXiv:1210.1833](https://arxiv.org/abs/1210.1833))
 - [4] Amendola L., et al., 2013, *Living Reviews in Relativity*, 16
 - [5] Anderson L., et al., 2012, *Monthly Notices of the Royal Astronomical Society*, 427, 3435–3467
 - [6] Angulo R. E., Pontzen A., 2016, *Monthly Notices of the Royal Astronomical Society: Letters*, 462, L1–L5
 - [7] Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, , 47, 481
 - [8] Asquith R., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 480, 1197
 - [9] Bagley M. B., et al., 2020, arXiv e-prints, p. [arXiv:2006.03602](https://arxiv.org/abs/2006.03602)
 - [10] Behroozi P. S., Wechsler R. H., Wu H.-Y., 2012, *The Astrophysical Journal*, 762, 109
 - [11] Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, *Astronomy & Astrophysics*, 564, A66
 - [12] Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *Astrophys. J.*, 345, 245
 - [13] Chuang C.-H., et al., 2019a, *Monthly Notices of the Royal Astronomical Society*, 487, 48
 - [14] Chuang C.-H., et al., 2019b, *Monthly Notices of the Royal Astronomical Society*, 487, 48–59
 - [15] Colbert J. W., et al., 2013, *Astrophys. J.*, 779, 34
 - [16] Collaboration D., et al., 2016, The DESI Experiment Part I: Science, Targeting, and Survey Design ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
 - [17] Croton D. J., et al., 2006, , 365, 11
 - [18] Croton D. J., et al., 2016, *The Astrophysical Journal Supplement Series*, 222, 22
 - [19] Daddi E., et al., 2007, *The Astrophysical Journal*, 670, 156–172
 - [20] Davidzon I., et al., 2017, , 605, A70
 - [21] Dawson K. S., et al., 2012, *The Astronomical Journal*, 145, 10
 - [22] Dawson K. S., et al., 2016, *The Astronomical Journal*, 151, 44
 - [23] De Lucia G., Blaizot J., 2007, , 375, 2
 - [24] De Lucia G., Kauffmann G., White S. D., 2004, *Monthly Notices of the Royal Astronomical Society*, 349, 1101
 - [25] Draine B. T., 2003, *Annual Review of Astronomy and Astrophysics*, 41, 241
 - [26] Drinkwater M. J., et al., 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 1429–1452

- [27] Eisenstein D., the SDSS-III collaboration 2016, Astronomers map a record-breaking 1.2 million galaxies to study the properties of dark energy, <https://www.sdss.org/press-releases/astronomers-map-a-record-breaking-1-2-million-galaxies-to-study-the-properties-of-dark-energy/>
- [28] Elbaz D., et al., 2011, *Astronomy & Astrophysics*, 533, A119
- [29] Favole G., et al., 2016, , 461, 3421
- [30] Favole G., et al., 2019, arXiv e-prints, p. [arXiv:1908.05626](https://arxiv.org/abs/1908.05626)
- [31] Groves B. A., Dopita M. A., Sutherland R. S., 2004, *The Astrophysical Journal Supplement Series*, 153, 75–91
- [32] Gruppioni C., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 3419
- [33] Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, , 343, 75
- [34] Izquierdo-Villalba D., Bonoli S., Spinoso D., Rosas-Guevara Y., Henriques B. M. B., Hernández-Monteagudo C., 2019, , 488, 609
- [35] Kennicutt Jr R. C., 1998, *Annual Review of Astronomy and Astrophysics*, 36, 189
- [36] Knebe A., et al., 2018a, *MNRAS*, 474, 5206
- [37] Knebe A., et al., 2018b, , 475, 2936
- [38] Kormendy J., Ho L. C., 2013, *Annual Review of Astronomy and Astrophysics*, 51, 511–653
- [39] Laureijs R., et al., 2011, Euclid Definition Study Report ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- [40] Lee J., et al., 2020, arXiv e-prints, p. [arXiv:2006.01039](https://arxiv.org/abs/2006.01039)
- [41] McConnell N. J., Ma C.-P., 2013, *The Astrophysical Journal*, 764, 184
- [42] Mo H., Mao S., White S. D., 1998, *Monthly Notices of the Royal Astronomical Society*, 295, 319
- [43] Nuza S. E., Sanchez A. G., Prada F., Klypin A., Schlegel D. J., Gottloeber S., Montero-Dorta A. D., Manera M. e. a., 2012, preprint, ([arXiv:1202.6057](https://arxiv.org/abs/1202.6057))
- [44] Orsi , Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 799–814
- [45] Osterbrock D. E., 1989, *Astrophysics of gaseous nebulae and active galactic nuclei*
- [46] Pozzetti L., et al., 2016, , 590, A3
- [47] Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 651–687
- [48] Spergel D., et al., 2013, Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report ([arXiv:1305.5422](https://arxiv.org/abs/1305.5422))
- [49] Spergel D., et al., 2015, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report ([arXiv:1503.03757](https://arxiv.org/abs/1503.03757))
- [50] Springel V., Yoshida N., White S. D., 2001, *New Astronomy*, 6, 79–117
- [51] Tremonti C. A., et al., 2004, *The Astrophysical Journal*, 613, 898–913
- [52] White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- [53] Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, , 435, 3500
- [54] Zhai Z., Benson A., Wang Y., Yepes G., Chuang C.-H., 2019, *Monthly Notices of the Royal Astronomical Society*, 490, 3667–3678
- [55] de Jong R. S., et al., 2012, *Ground-based and Airborne Instrumentation for Astronomy IV*

Appendices

A. ADDITIONAL VALIDATION PLOTS

In this Appendix we provide supplementary plots that further show the validity of the SAGE galaxies used throughout this study.

A.1. The Black Hole to Bulge Mass Relation

One of the most studied relations for galaxy statistics during the last decades has been the relation between the mass of the supermassive black hole (SMBH) present in the center of presumably almost all galaxies and the mass that characterizes the bulge region of such galaxies. By representing these two quantities against each other it has been found that most galaxies are distributed in such a way that a clear linear trend can be appreciated.

In the main panel of Fig.11 we have represented the mass of the central SMBH versus the mass of the bulge for the UNITSIM1-SAGE galaxies in an analogous way to how the data shown in Fig.4 is portrayed. In addition, in the bottom panel the variations between simulations is represented in the same way as we have been doing in Sec.III.

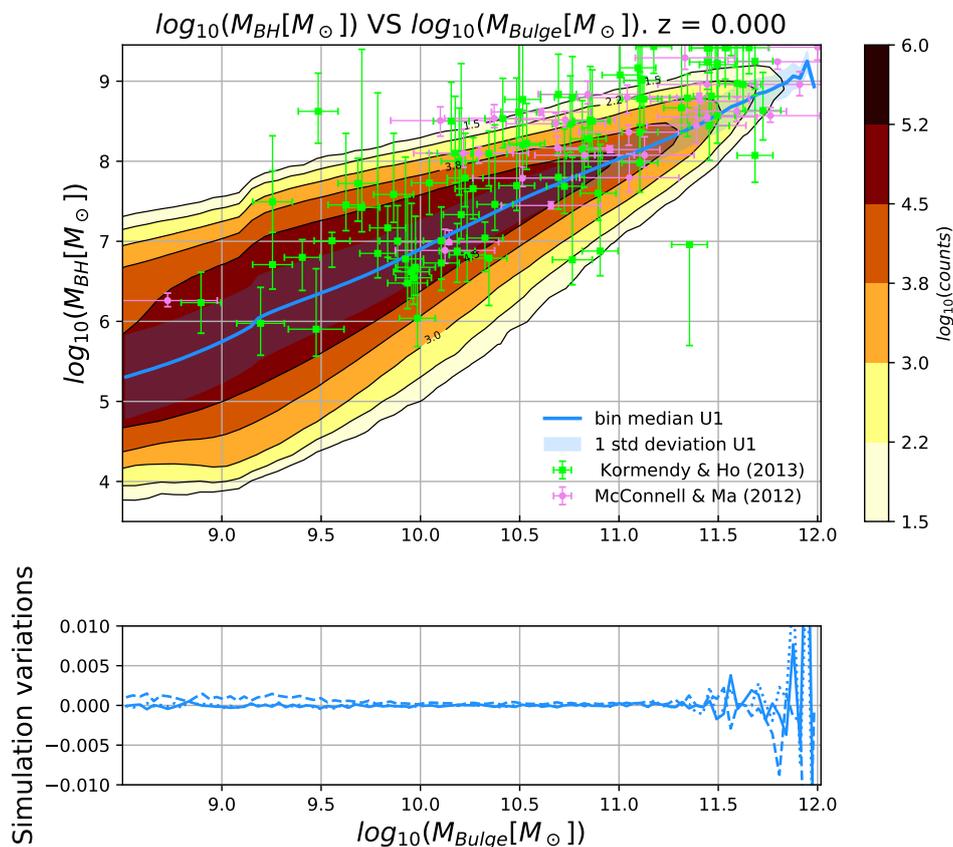


Figure 11: Black hole-bulge mass relation.

The main panel also shows observational data collected in Kormendy & Ho [38] and McConnell & Ma [41], which seems to be consistent with the results that we have obtained in our simulations. The results presented here are again also compatible with the results from Favole et al. [30], therefore, we are inclined to expect that our galaxies are capable of mimicking the correct behaviour for the redshifts of interest of Euclid.

A.2. The Cold Gas Fraction

Another property that must also be studied closely for our sample of simulated galaxies is the amount of cold gas that exists in each of them since the processes of star formation strongly depends upon the amount of cold gas available.

In order to compare the results of our simulations with observational data we present in Fig.12 the cold gas fraction (CGF), i.e. the ratio between the mass contained in the form of cold gas in a galaxy with respect to the stellar mass of that particular galaxy as a function of the stellar mass of such galaxy. As can be seen, this figure also shows the observational data for the CGF as a function of the stellar mass presented in Boselli et al. [11]. These values are in perfect agreement with the results of our simulations and, of course, also with previous results that can be found in Favole et al. [30].

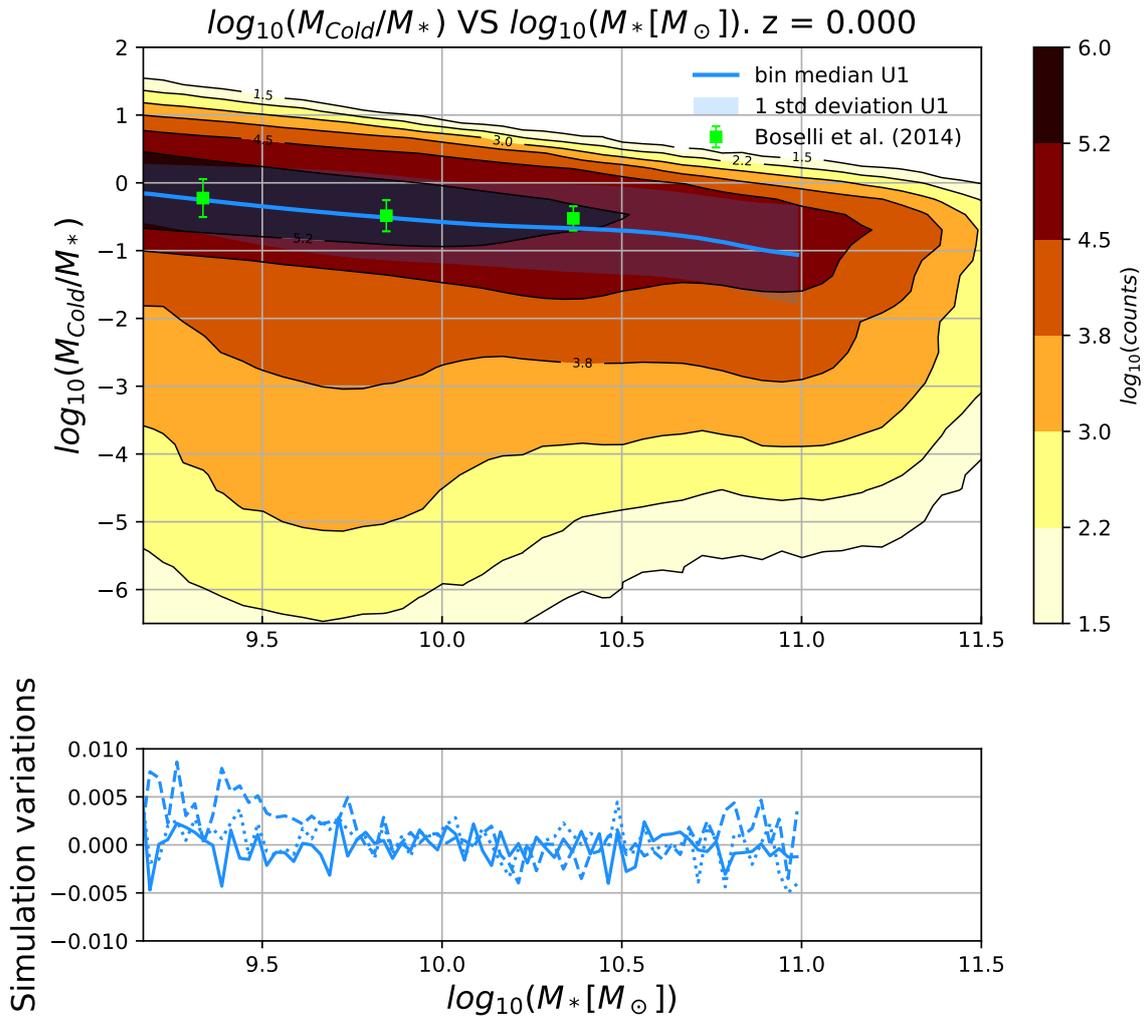


Figure 12: Cold gas fraction against stellar mass.

A.3. Stellar-to-Halo Mass Relation

The last property that we are going to study in this section is the relation between the stellar mass of the UNIT-SAGE galaxies and the mass corresponding to the halos that give shelter to these galaxies which have been produced by applying the ROCKSTAR code to the particle distributions at different redshifts from the UNIT simulations.

In the left panel of Fig.13 the stellar mass of the UNITSIM1-SAGE galaxies is represented versus the mass of the halos that contain such galaxies at redshift $z \approx 1$. The results are presented in an analogous way to previous sections and alongside the bin median for the stellar mass of the galaxies we have plotted the prediction for the relation between these two variables derived by Rodríguez-Puebla et al. [47].

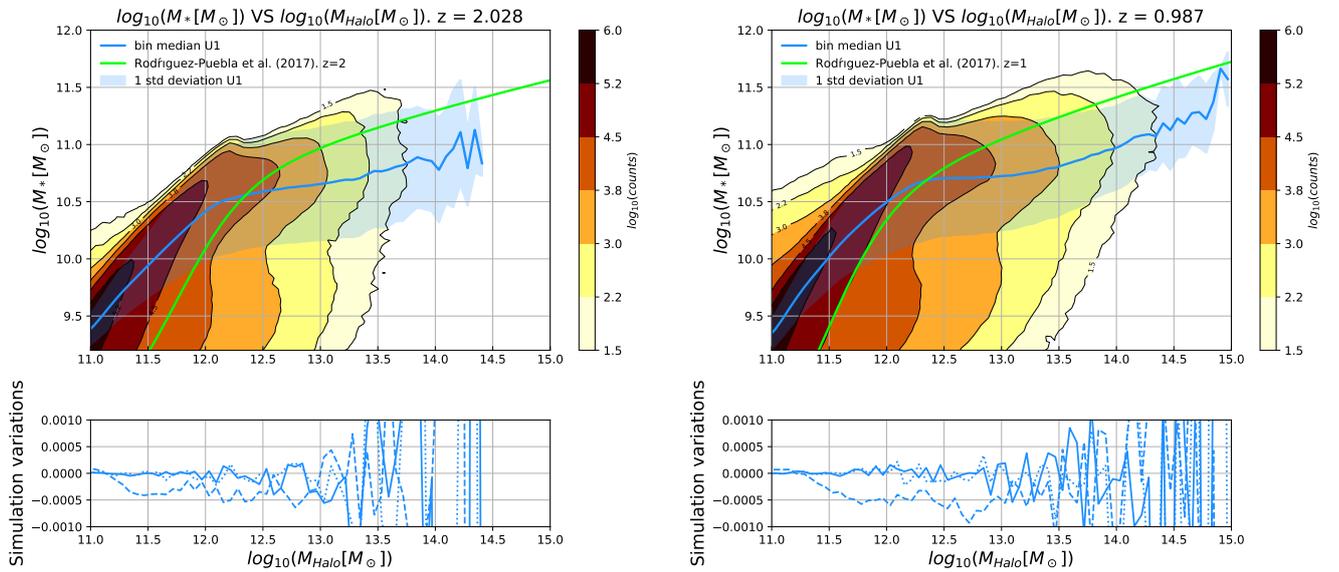


Figure 13: Stellar-to-halo mass relation.

In the right panel of Fig.13 the Stellar-to-Halo Mass relation is presented for redshift $z \approx 2$ (both from the simulations and from the Rodríguez-Puebla et al. [47] model).

As can be seen in both figures shown in this section, the results obtained for the simulated galaxies are similar to the prediction proposed for the Stellar-to-Halo Mass relation in Rodríguez-Puebla et al. [47] and seem to indicate, together with the rest of the results, that our sample of semi-analytical galaxies reasonably reproduces the characteristics and properties that would be expected to be observed for galaxies at redshifts $0.9 < z < 1.8$.