PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

FACULTAD DE FISICA
INSTITUTO DE ASTROFISICA

NEW INSIGHTS AND METHODS FOR THE CLUSTERING OF EMISSION LINE GALAXIES

BY ESTEBAN JIMÉNEZ

Tesis presentada a la Facultad de Física de la
Pontificia Universidad Católica de Chile, para
optar al grado académico de Magíster en Astrofísica

Advisor	:	Prof. Nélson Padilla (PUC Chile)
Co-advisor	:	Prof. Idit Zehavi (CWRU USA)
Correctors	:	Prof. Domenico Sapone (U. de Chile)
		Prof. Felipe Barrientos (PUC Chile)

July, 2019
Santiago, Chile

Abstract

We study the dependence of emission line galaxies (ELG) clustering in halo properties besides mass, an effect known as assembly bias (AB). This AB signature is scale-dependent in some ELG samples and may contribute with potential systematic effects on cosmology constraints. We test different implementations of the halo occupation distribution (HOD) model to construct mock galaxy catalogues that mimic ELG selected samples produced by a semi-analytical model (SAM). These catalogues are useful to understand the future observational samples. We select galaxies from two different SAMs, applied to the Millennium-WMAP7 and the MultiDark Plank 2 simulation. We use fixed number density galaxy samples selected according to stellar mass, star formation rate (SFR), and emission line luminosity from Hα, [OIII]λ5007, and [OII]$\lambda\lambda$3727 − 3729 emission, obtained from a photoionization code. We develop three different schemes to populate haloes with galaxies with increasing complexity, considering the scatter of the satellites HOD as an additional parameter in the modellings. The quality of the HOD modelling is determined by comparing the two-point correlation function (2PCF) of the HOD mocks and the SAM samples. The HOD modelling that uses the HOD of centrals and satellites separately, and assume a negative binomial distribution for the satellite HOD, produce the best clustering predictions. The scatter in the satellites HOD is a key consideration for HOD mock catalogues that mimic ELG selected samples in future galaxy surveys. The origin of the scale-dependent AB seems to be related to an environment selection effect rather than a physical reason. We found that galaxies in underdense regions and low gas metallicity, make the main contribution to the scale-dependent signature. We obtain that the peak of the Baryon Acoustic Oscillation (BAO) for some ELG samples is in lower scales than expected, and the β parameter is non-constant in large scales. Hence, this selection effect can introduce systematic in the inferred cosmology from ELGs.

Contents

Chapter 1

Introduction

In the current cosmological paradigm, the Universe is composed of a filamentary network of structures shaped by gravity. In this framework, dark matter haloes correspond to overdense regions that evolve by gravitational instability due to mergers and interactions with other haloes. Galaxy formation occurs inside haloes where baryons collapse in the gravitational potentials and the condensation of cold gas allows the formation of stars and the evolution of galaxies (White & Rees, 1978). As galaxy distribution is correlated with the mass distribution, accurate cosmology constraints can be inferred from galaxy surveys if a detailed description of the galaxy-halo connection is available. The evolution of dark matter haloes can be followed, to high accuracy, using N-body simulations which use a set of cosmological parameters as inputs. A complete description of the link of galaxies with their host haloes is difficult to obtain because of the complex physical processes that shape galaxy formation and evolution.

There are three major approaches to study this connection. One of them is the hydrodynamical simulation which resolves numerically the non-linear equations that describe the physical processes that impact the fate of baryons in dark matter haloes (e.g. Vogelsberger et al., 2014; Schaye et al., 2015). Despite the detailed treatment on baryon physics, these models are computationally expensive and cannot be run over the large volumes needed for cosmological studies. The second option is the Semi-analytical models (SAMs) that use simplified physical models of the processes that shape the evolution of baryons (Cole et al., 2000; Baugh, 2006; Benson, 2010; Somerville & Davé, 2015). The SAMs create galaxy catalogues for large volume dark matter simulations. The

specific recipes, used by these two physical models, can be tested by comparing the resulting predictions, as the abundance and clustering of galaxies, with large surveys data. This can provide useful constraints in the modellings as many physical processes are still poorly understood.

Another way to describe the galaxy population is with the halo occupation distribution (HOD) framework (Benson et al., 2000; Peacock & Smith, 2000; Scoccimarro et al., 2001; Yang et al., 2003). This is an empirical approach that provides a relation between the mass of haloes and the number of galaxies hosted by them. This is expressed as the probability distribution $P(N|M_h)$ that a halo of virial mass M_h hosts N galaxies which satisfy some selection criteria. This approach provides insight into the halo-galaxy connection and can be used to study galaxy clustering (Berlind & Weinberg, 2002; Zheng et al., 2005; Conroy et al., 2006; Zehavi et al., 2011; Wechsler & Tinker, 2018). Furthermore, the HOD parameters can be tuned in detail because they only aim to reproduce a limited set of observables such as the galaxy number density and clustering. Thus, HOD modelling is one of the most efficient ways to populate very large volumes or to produce many realizations required for, e.g., estimating covariance matrices using mock galaxy catalogues (e.g. Norberg et al., 2009; Manera et al., 2013). These mock catalogues can then be used to test and develop new algorithms that will be used for the next generation of surveys.

The study of star forming emission line galaxies (ELGs) has gained interest over the last decade as they will be targeted by surveys such as Euclid and the Dark Energy Spectroscopic Instrument (DESI) surveys (Laureijs et al., 2011; DESI Collaboration et al., 2016). The luminosity of an emission line depends on a number of factors, including the star formation rate (SFR), gas metallicity and the conditions in the HII regions (e.g. Orsi et al., 2014). Even though ELGs samples are related to star formation, they are not the same as SFR-selected samples. Still, a similar HOD approach can be used to study both galaxy populations (Geach et al., 2012; Cochrane et al., 2017; Cochrane & Best, 2018). In particular, the shape of the HOD in SFR selected samples is more complex than the case of the more widely studied stellar mass selected samples (e.g. Contreras et al., 2013; Gonzalez-Perez et al., 2018). For example, the occupation function of central galaxies in ELG samples does not follow the canonical step-like form. Accurate modelling of the HOD will provide the more realistic mock catalogues needed for the analysis of future observational samples.

Using ELGs as tracers of the underlying density field opens new options to tighten the current cosmological parameters (Delubac et al., 2017; Orsi & Angulo, 2018; Grasshorn Gebhardt et al.,

2019). For example, current and coming surveys aim to detect high precision measurements of Baryon Acoustic Oscillations (BAO) in the power spectra of ELGs. The constraints on cosmological parameters can be obtained combining the measures of the BAO scale, with the Cosmic Microwave Background (CMB) and Supernova information. In the first order prediction, the linear theory of density perturbations shows that the clustering of galaxies and haloes are related by a constant bias factor at large scales (see Mo et al. (2010)). The bias is scale-dependent for intermediate scales (i.e r $\lesssim$ 30 Mpc h^{-1}) (Blanton et al., 1999). But, if bias is non-constant for large scales, then environment effects may be encoded in the large scale structure of galaxies. This effect can potentially impact the BAO scale, and introduce systematic when inferring cosmological parameters.

Here, we use the HOD formalism to test three different ways to populate dark matter haloes with galaxies. The prescriptions of these models aim to replicate as accurately as possible the target galaxy populations of a SAM sample. The comparison between the galaxy population in the mock catalogues and SAM samples is done via the analysis of their two-point correlation function (2PCF), which is related to the power spectra of density fluctuations and is sensitive to cosmology (e.g. DeRose et al., 2019). We also include the scatter of the HOD of satellites in our modelling, and quantify the impact of using this additional parameter on the clustering.

Additionally, we study the clustering of different ELG samples at large scales. We compute the impact on clustering from additional properties besides halo mass, an effect that is known as assembly bias (Gao et al., 2005; Croton et al., 2007). This signature shows a constant evolution with scale-separation either in stellar mass and SFR selected samples, (Zehavi et al., 2018; Contreras et al., 2019; Zehavi et al., 2019). We study the large-scale structure of galaxies computing the local number densities of their host haloes, and galaxy properties as gas metallicity that can be related with features in the clustering of ELGs. In particular, we examine a scale-dependent assembly bias signature, found in some ELG samples, and we relate this with a non-constant large scale bias. We also discuss possible origins of this signal and the potential impacts on the BAO scale and in a cosmological parameter.

The outline of this thesis is divided into two big chapters. In Chapter 2, we present a method that produces accurate mock galaxy catalogues that mimic SFR-selected samples (related to ELGs) of the next generation of surveys. This chapter, and part of the introduction and conclusion are

verbatim from Jiménez et al. (2019). In Chapter 3, we present new insights about the clustering of emission line galaxies, which can potentially impact the inferred cosmological parameters from future ELG data.

Chapter 2

HODs for accurate clustering predictions[1]

In this chapter, we present an extension of the HOD method to construct mock catalogues. We obtain the galaxy catalogues from a semi-analytical model which is applied in the Millennium simulation. In § 2, we show the description of the galaxy formation model and the definition of our galaxy samples. In § 3, we measure the 2PCFs and the HODs of these galaxy samples. In § 4, we present our HOD models and the parametrization of the scatter of satellite's HOD. Finally, we show the accuracy of the HOD models and the impact of the HOD scatter in § 5.

2.1 Simulation data

In this section we give a brief overview of the galaxy formation model used (§ 2.1) and the N-body simulation in which it is implemented (§ 2.2).

2.1.1 Galaxy formation model

A galaxy formation model needs to take into account a variety of physical processes such as radiative cooling of gas; AGN, supernovae and photoionisation feedback; chemical evolution;

[1]This chapter is a verbatim from the paper *Extensions to the halo occupation distribution model for more accurate clustering predictions*, MNRAS, submitted; arXiv:1906.04298 by E. Jiménez, S. Contreras, N. Padilla, I. Zehavi, C.M. Baugh and V. Gonzalez-Perez.

star formation; disc instabilities; collapse and merging of dark matter haloes; and galaxy mergers. These affect the fate of baryons in haloes which lead to the formation and evolution of galaxies. Several physical processes such as star formation and gas cooling are not fully understood due to their complexity. As a consequence, a set of free parameters are used in the equations that model these processes. These free parameters are tuned in order to reproduce observations such as the luminosity functions, colours and the distribution of morphological types. In this context, different SAMs usually have their own implementations to model these physical processes, predicting different galaxy populations. Here we use the outputs at $z = 0$ from the SAM of Guo et al. (2013) (hereafter G13) which is a version of the `L-GALAXIES` code from the Munich group (De Lucia et al., 2004; Croton et al., 2006; De Lucia & Blaizot, 2007; Guo et al., 2011; Henriques et al., 2013). The outputs are publicly available from the Millennium Archive[2].

The samples used here are defined according to three different number densities where we rank the galaxies in the SAM by their stellar mass or SFR in a decreasing way (hereafter the SAM samples). These samples are useful in order to compare with observational catalogues with similar space densities. Table 2.1 shows the three number densities and the cuts in stellar mass and SFR used in each case.

2.1.2 The Millennium simulation

The distribution of dark matter haloes used in this work is drawn from the Millennium-WMAP7 simulation Guo et al. (2013) which is identical with the Millennium Simulation Springel et al. (2005), but with updated cosmological parameters that match the results from the WMAP7 observations. This version assumes a flat ΛCDM universe considering $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.704$ and $\sigma_8 = 0.81$. The simulation was carried out in a box-size of 500 h^{-1}Mpc following 2160^3 particles of mass $9.31 \times 10^8 h^{-1}\text{M}_\odot$. The run produced 61 simulation snapshots from $z = 50$ up to $z = 0$. G13 use a friends-of-friends group finding algorithm (`FOF`) to identify dark matter haloes in each snapshot (Davis et al., 1985) and then run `SUBFIND` to identify the subhaloes (Springel et al., 2001). Halo merger trees are constructed for each output and track the evolution of haloes through cosmic time. These trees are the starting points for the SAM.

[2] `http://gavo.mpa-garching.mpg.de/Millennium/`

$n/h^3\mathrm{Mpc}^{-3}$	$\mathrm{M}^*_{\mathrm{min}}/h^{-1}\mathrm{M}_\odot$	$\mathrm{SFR}_{\mathrm{min}}/\mathrm{yr}^{-1}\mathrm{M}_\odot$
$10^{-3.0}$	5.95×10^{10}	5.25
$10^{-2.5}$	3.38×10^{10}	2.53
$10^{-2.0}$	1.25×10^{10}	0.70

Table 2.1: The first column shows the abundance of galaxies in the three density samples used here. The second and third columns show the cuts applied to G13 galaxies in stellar mass and star formation rate, respectively, to achieve these abundances.

2.2 Characterization of the SAM galaxy samples

This section introduces the statistics used to characterize the distribution of galaxies, starting with the measurement of the correlation function (§ 3.1), the form of the HOD predicted by the SAM (§ 3.2) and the scatter in the HOD (§ 3.3).

2.2.1 Clustering measurement: two-point galaxy correlation function

The spatial two-point correlation function, $\xi(r)$, measures the excess probability of finding a pair of galaxies at a given separation with respect to a random distribution. We compute the 2PCF of the galaxy samples with the `Corrfunc` code (Sinha & Garrison, 2017).

Fig. 2.1 shows the 2PCF of the stellar mass (top) and SFR (bottom) selected samples for the different space densities. For the former, the amplitude of the clustering increases with decreasing number density, as we consider more massive galaxies. The impact of the inclusion of these massive galaxies is stronger at small scales and is weaker at large scales. In contrast, for the SFR selected galaxies the amplitude of the 2PCF for the different samples remains largely unchanged except for small scales where the satellite-satellite pairs make an important contribution to the clustering amplitude. For both selections, the satellite fraction increases with increasing number density.

In the 2PCF, we can distinguish between the contribution from galaxy pairs in the same halo and from different haloes. The former are the main contributors to the amplitude of the 2PCF on small scales, namely the one-halo term which dominates up to $\sim 1\ h^{-1}$Mpc, while galaxy pairs between different haloes contribute mostly to the two-halo term which determines the clustering on large scales. In this regime the total number of galaxies in the halo, regardless of whether they are satellites or the central, drives the amplitude of the clustering, acting as a weighting for the bias

of each halo in computing an overall "effective" bias for the sample (see e.g. Baugh et al. 1999). The one-halo term is sensitive not only to the number of satellites, but also depends on their spatial distribution.

2.2.2 The halo occupation function predicted by the SAMs

The galaxy populations in the SAMs depend on the choices adopted for the modelling of the baryonic processes. Hence, depending on the SAM employed, different galaxy catalogues with different luminosity functions, stellar mass functions or correlation functions can be obtained for the same dark matter simulation. For example, Contreras et al. (2013) studied the effects on the clustering predicted from different SAMs and found some differences particularly in galaxy samples selected by SFR and cold gas mass. Moreover, they show that the shapes of the HODs are model-dependent which reflects the differences in the implementation of physical processes in each SAM. For example, the specific modelling of dynamical friction affects the satellite population in SAMs. Here we are not interested in the detailed shape of the HOD predicted by a particular SAM, but on how best to use the occupation functions to populate dark matter haloes with galaxies to produce a similar spatial distribution to that resulting from a SAM.

The HOD is usually broken down into the contribution from central and satellite galaxies. Fig. 2.2 shows these two components for stellar mass and SFR selected samples with the same number density for the G13 SAM. Here, each HOD is computed in bins of width 0.08 dex in the logarithm of the halo mass where the position of each $\langle N \rangle$ value is plotted at the median value within each bin. The striking difference in the shape of the HOD of centrals between the two selections is due to the different galaxies that are included. Massive centrals tend to be red galaxies hosted by massive dark haloes. Such centrals are included in stellar mass selected samples but not when selecting by SFR. The galaxies in the SFR samples correspond mainly to blue star-forming galaxies excluding luminous red galaxies with high stellar mass but low SFR. It is noteworthy that the fraction of halos that contain a central passing the SFR selection never reaches unity for the sample plotted in Fig. 2.2. These features of the HOD of SFR galaxies have been noted in SAMs before (e.g. Contreras et al., 2013, 2019; Gonzalez-Perez et al., 2018) and inferred for blue galaxies in the SDSS (e.g. Zehavi et al., 2011). This shows that a significant number of haloes in the SFR selected samples do not host a central as their SFR is below the threshold. The same situation is

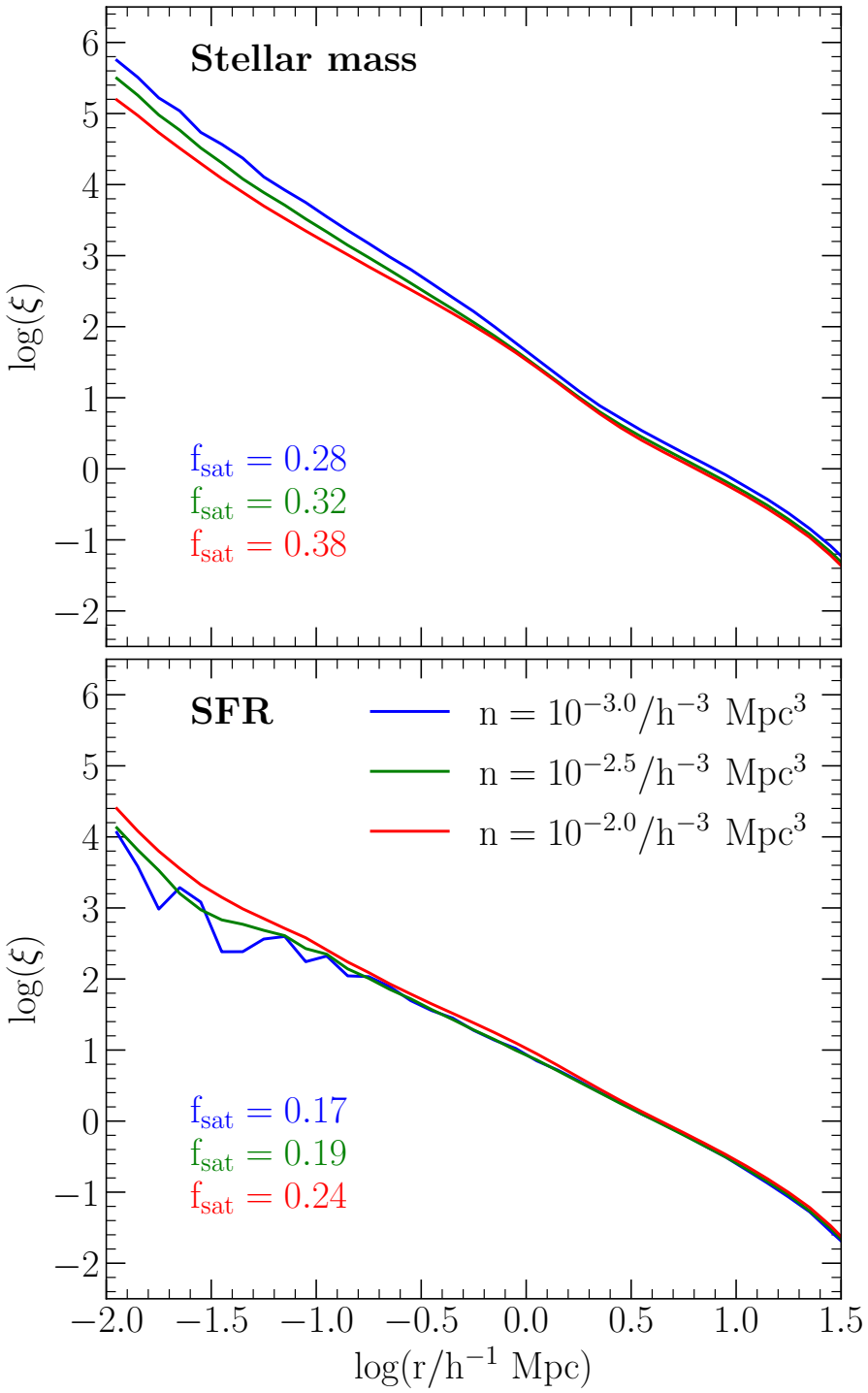

Figure 2.1: Two-point correlation functions ($\xi(r)$) of different galaxy samples from G13 and defined in Table 2.1. (Top) stellar mass and SFR selected samples (bottom). Colors indicate each sample as labeled in the bottom panel. The fraction of satellites in each sample is shown in both panels, with the color indicating the sample number density.

found in the other number density samples. Note that, in observational samples, the ranking of galaxies in order of their emission line luminosity may not correspond to the ranking in SFR due to dust attenuation, which means that the highest SFR galaxies may not necessarily have the brightest emission lines.

We estimate the uncertainties of the HOD values using jackknife resampling (Norberg et al., 2009), dividing the simulation volume into 10 slices. We use the position of the centre of the potential of haloes to classify the galaxies within each halo. The resulting errors are shown as the shaded regions in Fig. 2.2, and they are negligible for all halo masses except at the high mass end and for the HOD of centrals selected by SFR.

Because of the simple relation between halo mass and occupation number, the HOD represents a useful approach for the construction of mock galaxy catalogues. Here we have described the first moment of the HOD, the main ingredient to building-mocks recipes. Nevertheless, it is important to also consider the second moment, i.e the dispersion in the HOD of satellites.

2.2.3 The predicted dispersion in the halo occupation number

When the simplest HOD approach is used to build mock catalogues, the mean of the distribution is the main parameter. Central galaxies are assumed to follow a nearest-integer distribution where the mean $\langle N_{cen} \rangle$ is between zero and one. For satellites, a Poisson distribution with mean $\langle N_{sat} \rangle$ is the most widely assumed distribution (e.g Kravtsov et al., 2004; Zheng et al., 2005).

In G13, satellites are classified as type-1 if they are hosted by a resolved subhalo, and type-2 or orphans if the subhalo has been destroyed by tidal effects and is no longer identified. Boylan-Kolchin et al. (2010) found that the number of low mass subhaloes in main haloes in the Millennium-II Simulation (Boylan-Kolchin et al., 2009) is well described by a negative binomial distribution which corresponds to a super-Poissonian statistic as its scatter is larger than a Poisson distribution. This suggests that the type-1 satellite population can also be described by this distribution. Based on the outputs from the SAM presented in Jiang & van den Bosch (2016), and using the Bolshoi (Klypin et al., 2011) and MultiDark (Prada et al., 2012) simulations, Jiang & van den Bosch (2017) showed that ignoring this non-Poissonity in the HOD of subhaloes results in systematic errors in the predicted clustering of galaxies. Here we extend the application of the negative binomial distribution by checking whether the HOD of G13 satellites, that is including type-1 and type-2, is

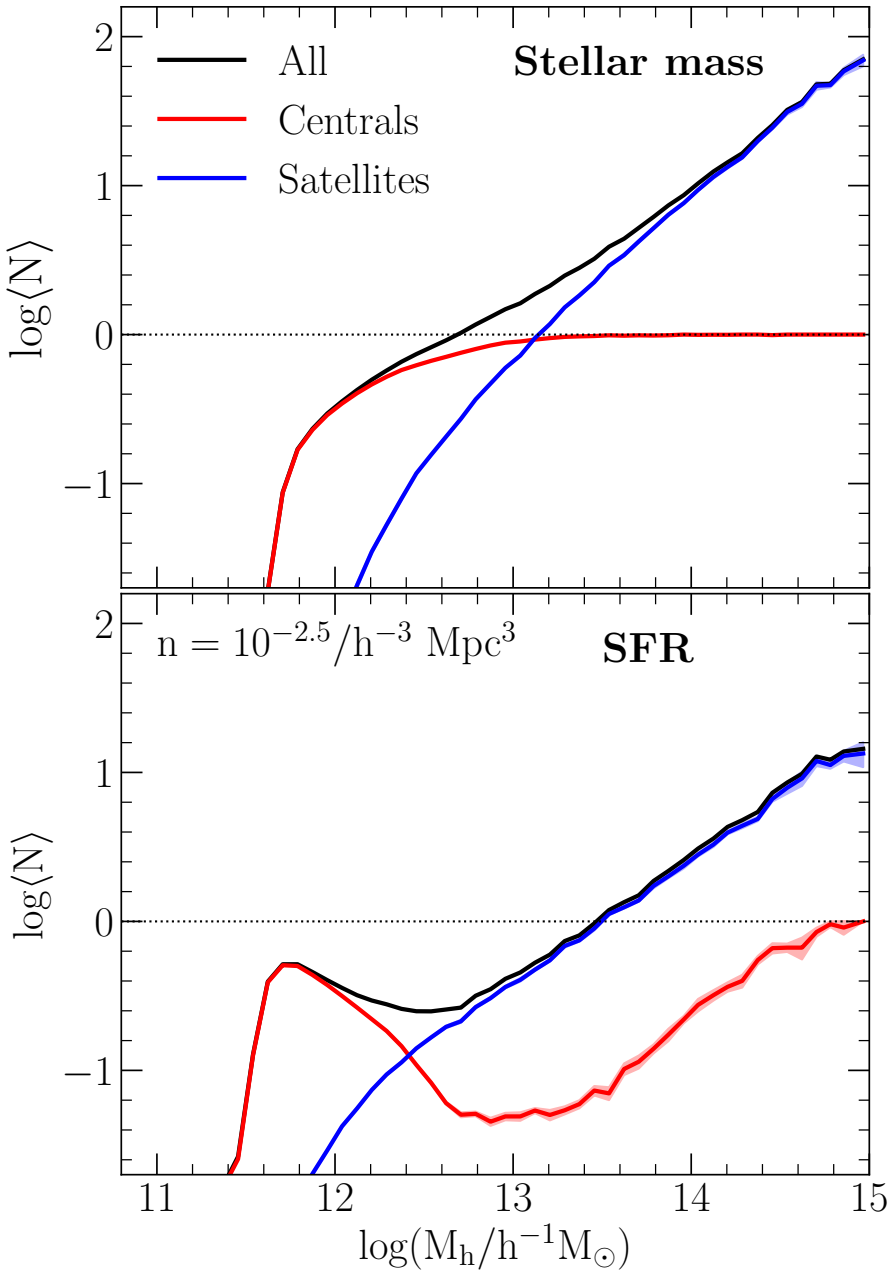

Figure 2.2: The HOD predicted by G13 for stellar mass (top) and SFR selected samples (bottom), for a number density of $10^{-2.5}h^3\text{Mpc}^{-3}$. Black lines show the HOD for the full sample and red and blue indicate the HOD for central and satellite galaxies, respectively. The red and blue shaded regions represent jackknife errors calculated using 10 subsamples. The horizontal black dotted line shows an average occupation value of unity.

well described by this statistic. We expect that the HOD scatter is model-dependent because of the different treatments of dynamical friction. Moreover, as some galaxy properties, such as SFR and stellar mass, have a model-dependent scatter it is reasonable to assume the same for HODs. For example, Guo et al. (2016) showed that different galaxy formation models do not have the same dispersion in the stellar mass-halo mass relation. Therefore our results are specific to the G13 model. It is likely that a different SAM would require an adjustment to to the value of β to describe the scatter of the satellite HOD. Nevertheless we expect our general results to hold for any SAM, and that the satellite distribution displays more scatter than Poisson.

The Poisson and negative binomial distributions differ in their shapes so it is useful to parametrise the departure from the Poisson scatter. We use the parameter β (defined below) to denote this departure. For a Poisson distribution the variance is given by the mean value of the random variable, namely $\langle \mathrm{N_{sat}} \rangle$, with the standard deviation given by $\sigma = \sqrt{\langle \mathrm{N_{sat}} \rangle}$. The negative binomial distribution has the same mean as the Poisson distribution, but a larger scatter which can be expressed as

$$\sigma_{\mathrm{NB}} = \sigma + \beta\sigma. \tag{2.1}$$

where $0 < \beta < 1$. Then, β indicates the fractional change in the variance with respect to the Poisson standard deviation σ. Under this definition, when $\beta = 0$ the distribution is Poissonian and if $\beta = 1$ the standard deviation is twice that from a Poisson distribution.

The probability function of the negative binomial distribution is given by

$$P(N\,|r, p\,) = \frac{\Gamma(N + r)}{\Gamma(r)\Gamma(N + 1)} p^r (1 - p)^N. \tag{2.2}$$

Here $\Gamma(x) = (x + 1)!$ is the gamma function. The parameters r and p are determined by the first moment $\langle N \rangle$ and second moment σ^2 of the distribution,

$$p = \frac{\langle N \rangle}{\sigma^2_{\mathrm{NB}}}, \quad r = \frac{\langle N \rangle^2}{\sigma^2_{\mathrm{NB}} - \langle N \rangle}. \tag{2.3}$$

Thus, we can control the width of the negative binomial distribution through the parameter β and compute the value of σ^2_{NB}.

2.3 Generating HOD mock catalogues

We now describe the procedure followed to build HOD mock galaxy catalogues using the HODs of the SAM samples. Section 2.3.1 presents the three methods we use to populate haloes with galaxies. In Section 2.3.2 we specify the treatment of the scatter in the HOD of satellites. Section 2.3.3 explains how we impose a standard Navarro-Frenk-White (NFW) density profile for satellites. Section 2.3.4 presents the impact of assembly bias in the SAM samples and explains why it must be removed from the SAM in order to compare with the HOD mock catalogues. Finally in Section 4.5 we discuss the treatment of the radial distribution of satellite galaxies within haloes.

2.3.1 The HOD models used to build mocks

We test three different HOD schemes of increasing complexity. This helps us to understand the level of complexity needed to obtain accurate clustering predictions. Each model uses occupation functions obtained from linear interpolations of the HOD values in each bin, rather than fitting a parametric form to the SAM HOD. The distribution of galaxies can be nearest-integer (centrals only) and Poisson or negative binomial (satellites).

1-HOD: The 1-HOD model builds mock catalogues using the HOD of all galaxies from the SAM sample (black solid lines in Fig. 2.2) including both centrals and satellites. The model assumes either a Poisson or negative binomial distribution for the occupation number. We adopt a Monte Carlo approach to obtain the final number of galaxies.

This approach does not distinguish between centrals and satellites. If the model predicts that $N \geq 1$ we assume that this halo hosts a central and $N_{sat} = N - 1$. Because of this, the number of centrals and satellites in the 1-HOD mock catalogues can be notably different with respect to the SAM samples where there are haloes with satellites but no central. Moreover, the HODs of these two separate components in the mock catalogues are completely different with respect to the HODs of the SAM samples. However, the total number of galaxies in these mock catalogues is essentially the same as in the SAM samples.

2-HOD: The 2-HOD model uses the HOD of centrals and satellites separately, i.e, the red and blue solid lines in Fig. 2.2, respectively. Thus, a particular distribution can be assumed for each component and the modelling is done independently for each one. For centrals, we use the

nearest-integer distribution and for satellites the Poisson or negative binomial distribution.

This scheme predicts practically the same number of central and satellites as the SAM samples. Note that in a non-negligible number of realizations it is possible to get haloes without a central. This is more likely for haloes with masses for which $\langle N_{cen} \rangle < 1$ which is more frequently found in SFR-selected samples.

4-HOD: This model contains more information about the galaxy population of the SAM sample than the 2-HOD model. The 4-HOD requires us to store the number of haloes that host a central (N_{cen}) and the number of haloes that do not host a central (N_{nocen}) as a function of halo mass. Under this definition the total number of haloes in the volume is the sum of both quantities. Furthermore, the 4-HOD also needs knowledge of the number of satellites in haloes with a central (N_{sat_cen}) and without a central (N_{sat_nocen}). Thus, the total number of satellites is the sum of these two quantities. With these definitions, we build new HODs for satellites that take into account the population of centrals in the SAM samples. The SAM samples contain haloes with satellites but no centrals. This is more common in SFR selected samples. Indeed the HOD of centrals in these samples indicates that a large number of haloes do not host a central (see Fig 2.2), and the 4-HOD takes this feature into account.

We then define the satellite occupation functions conditioned on whether or not haloes host a central. With the four quantities explained above, we can define the conditional HODs,

$$\langle N_{sat\_cen}(M_h) \rangle = \frac{N_{sat\_cen}}{N_{cen}}(M_h) \tag{2.4}$$

$$\langle N_{sat\_nocen}(M_h) \rangle = \frac{N_{sat\_nocen}}{N_{nocen}}(M_h) \tag{2.5}$$

Fig. 2.3 shows the conditional HODs where the main differences are observed at low halo masses. Even though the ratio between these two HODs is close to unity, it is the galaxies hosted by these haloes ($\approx 10^{12} h^{-1} M_\odot$) that dominate the amplitude of clustering. The conditional HODs are well fitted by a negative binomial distribution, including the HODs of the other number density samples. The 4-HOD method uses a Monte Carlo approach to decide if a halo hosts a central galaxy. Depending on this outcome, one of the two conditional HODs is then chosen to obtain the number of satellites.

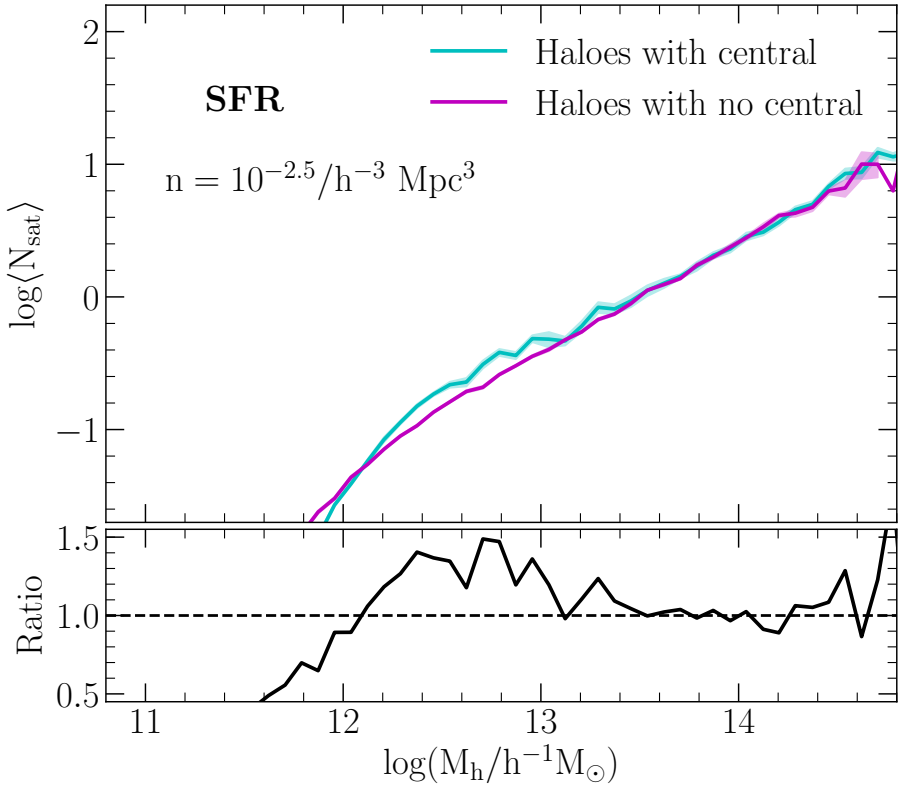

Figure 2.3: Conditional HODs from the 4-HOD method for an SFR selected sample with number density $10^{-2.5}h^3\,\mathrm{Mpc}^{-3}$. Top: Average number of satellites in haloes with a central (cyan) and without a central (magenta). Bottom: Ratio of the two HODs shown in the upper panel. Shaded regions represent jackknife errors calculated using 10 subsamples.

2.3.2 Treatment of scatter in the HOD of satellites

A Poisson distribution is fully described by its first moment. In the case of satellites this is $\langle N_{sat} \rangle$. If the distribution of the number of satellites follows instead a negative binomial distribution, an additional parameter β is needed which specifies the increase in the scatter with respect to a Poisson distribution (see Eq. 1). We fix the β value so that we reproduce as closely as possible the scatter of the HOD of satellites in a given SAM sample. Fig. 2.4 shows the scatter of the HOD of satellites in SAMs and 2-HOD mock catalogues for two illustrative β values in a SFR selected sample. This shows that a small but non-zero β is required to reproduce the HOD scatter of the SAM sample. The same is found for the other number density samples, and for the conditional HODs. We do not perform this analysis for the 1-HOD model as satellites are not treated independently in this case.

It is not possible to replicate the HOD scatter in SAMs more closely as this would require β to be a function of mass. Instead, we assume a constant scatter for the HOD of satellites by using the same β for all halo mass bins. The accuracy of the β values used are judged by checking the quality of the resulting mocks via comparison of their 2PCFs with the clustering of the shuffled-NFW samples (see § 4.5 below for the definition of this catalogue). We show in Section 2.4, that when the scatter of the SAM and HOD mocks are matched up to $M_h \lesssim 10^{13.5} M_{\odot} h^{-1}$ ($\beta = 0.05$ for SFR-selected samples), we obtain the most accurate clustering predictions. In contrast, using larger values for β worsens the predictions (as does using $\beta = 0$, which corresponds to Poisson scatter).

The satellite HOD is well described by the negative binomial distribution for a wide range of halo masses. Fig. 2.5 shows the satellite PDF in a particular mass bin for a stellar mass and a SFR selected sample. We show negative binomial distributions defined by $\beta = 0.08$ and $\beta = 0.05$. In order to compute the satellite distributions, we split satellites according to whether or not their haloes host a central galaxy, which is relevant for the 4-HOD model. The satellite distribution matches with the negative binomial when most of haloes in the bin are included. A similar close match is found when comparing with Poisson distributions ($\beta = 0$). Note that in the SFR selection case most of the haloes do not host a central galaxy, as the HOD of centrals in that bin suggests. The opposite behaviour is observed when selecting by stellar mass.

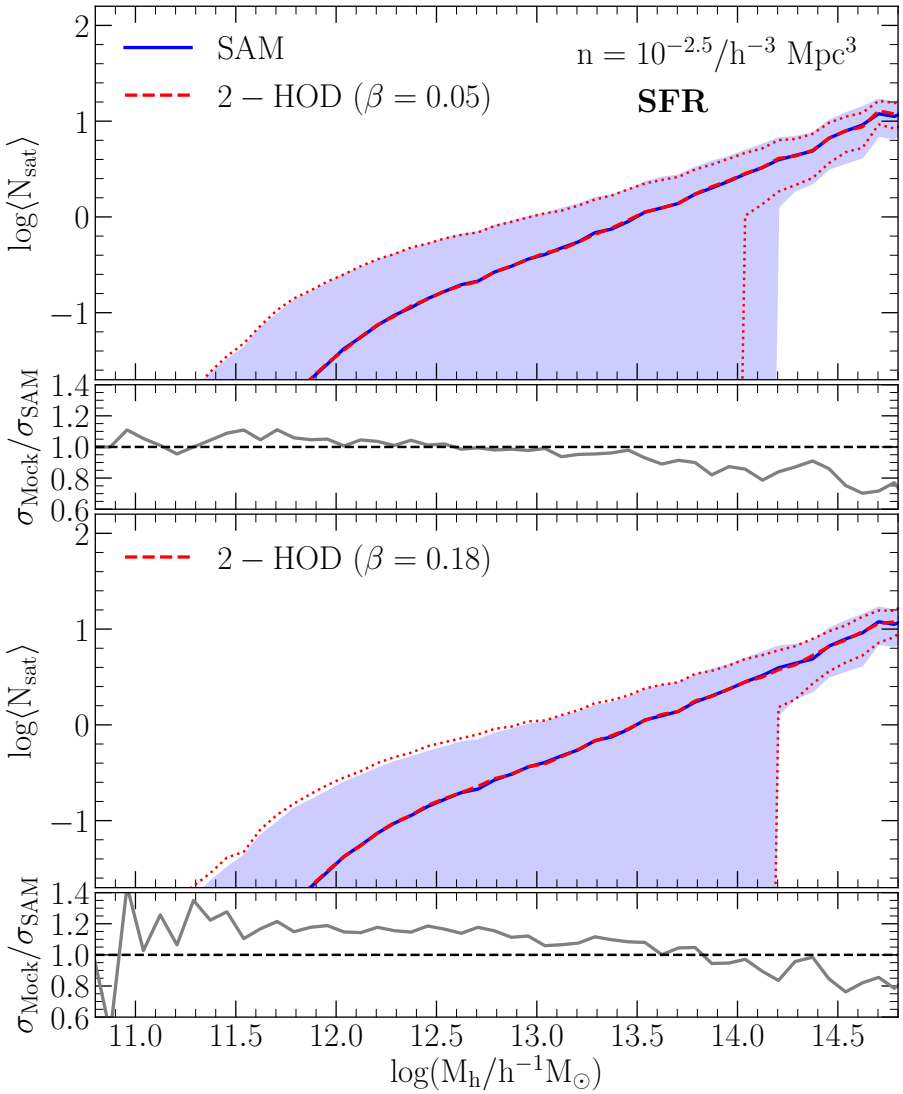

Figure 2.4: The HOD of satellites in a SAM sample (dashed blue) and a 2-HOD mock catalogue (solid red) contrasting two values of the parameter β that controls the scatter (see Eq. 1): $\beta = 0.05$ (top) and $\beta = 0.18$ (bottom). The shaded regions show the HOD scatter and the red dotted lines correspond to the scatter in the HOD mocks. The subpanels show the ratios between the HOD scatter of the mocks and SAM sample. Note that it is not possible to visually distinguish a Poisson scatter from the β scaled versions plotted in the main panels, but this choice would lead to a larger ratio of variances than the range plotted in the lower subpanels.

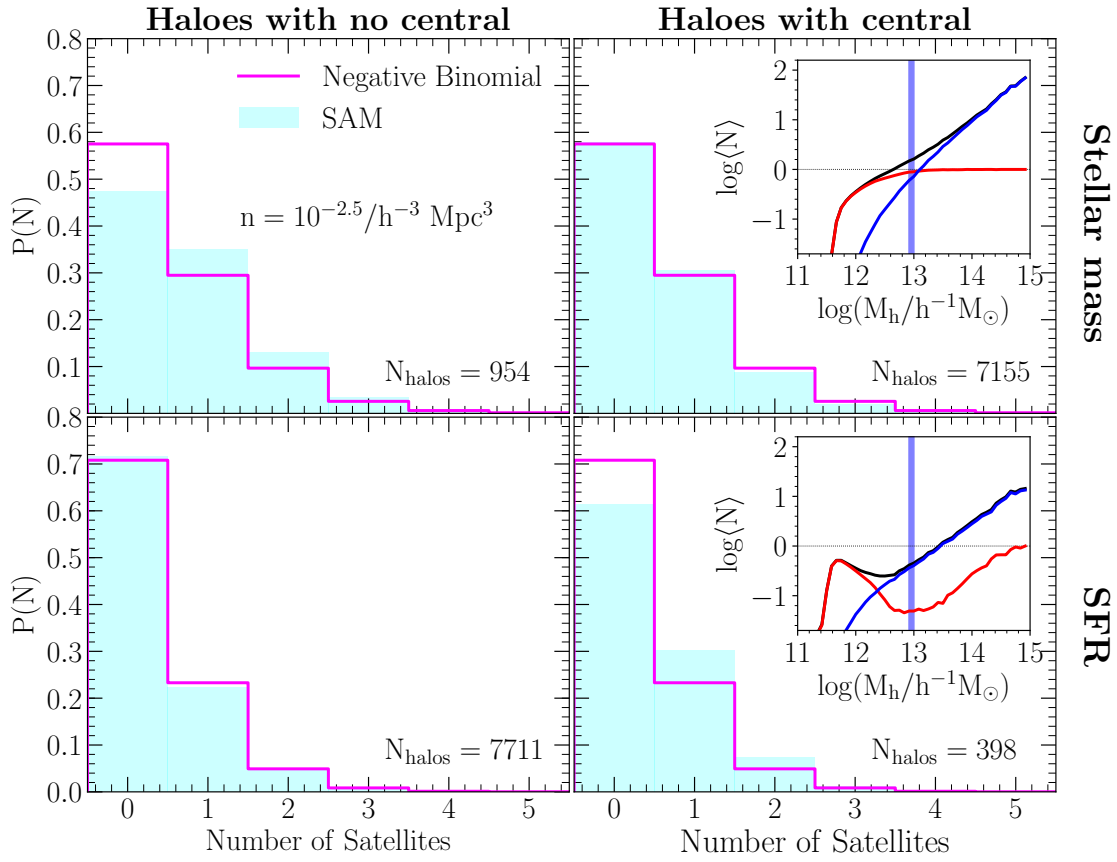

Figure 2.5: Probability distributions of satellites (cyan histograms) that are hosted by haloes with masses within the blue shaded (vertical) mass range of the HODs shown in the insets. The galaxies are selected by stellar mass (top) and SFR (bottom), with a number density of $10^{-2.5}h^3\mathrm{Mpc}^{-3}$. Left: Haloes without centrals (i.e its central did not enter to the cut). Right Same as left panel but considering haloes with centrals. Note the high probability to find haloes that do not host centrals in the SFR sample, which is expected by the low value of $\langle N_{cen}\rangle$ in the mass range analyzed and as is shown in the inset. The distributions are well described by negative binomial distributions (magenta) in the cases when most of haloes in the bin are included (i.e top right panel and bottom left panel). The negative binomial distributions shown here are obtained using a scatter that is 5% larger than that from a Poisson distribution in the SFR selected sample, and 8% larger in the stellar mass selection case

.

2.3.3 The radial distribution of satellite galaxies in halos

The number of satellites in the mock catalogues is obtained from the adopted HOD model (see Sec 2.3.1). Their positions in haloes are set according the standard NFW density profile (Navarro et al., 1996) which requires two parameters, the concentration and scale radius. The former depends on halo mass and the latter is a function of the virial radius. For simplicity, we assume that all haloes in the simulation volume have the same concentration parameter $c = 13.98$ which corresponds to the concentration of a halo at redshift z = 0 and mass $M_h = 10^{12.5}\mathrm{M}_{\odot}h^{-1}$. We do not use a more realistic model for concentration as we are interested in comparing the HOD models rather than obtain a realistic redistribution of satellites. We impose that the maximum distance from a satellite to the halo center is two virial radii which depends on the halo mass. This defines the NFW mass profile used to obtain the satellite distances by a Monte Carlo approach. We modify the SAM output to impose a similar satellite distribution as described below (§ 4.5).

2.3.4 Removing assembly bias from the SAM output

In order to determine the best methodology to produce HOD mock catalogues, we aim to compare them with the clustering of the original SAM samples via their 2PCF. Before making this comparison it is necessary to remove assembly bias from the SAM samples.

The clustering of dark matter haloes depends on additional properties besides mass. For example, Gao et al. (2005) showed that the clustering of low mass haloes depends on their formation redshift and other works have found dependencies on concentration and subhalo occupation number (e.g. Wechsler et al., 2006) among other secondary properties. This additional contribution to the clustering is commonly known as assembly bias and potentially changes the galaxy clustering amplitude on large scales.

The standard HOD approach considers only halo mass as the variable regulating the galaxy population. SAMs include assembly bias because they follow the evolution of baryons in halo merger histories that are shaped by the large-scale environment in the N-body simulation. Namely, SAMs include a dependence on secondary halo properties as these affect the halo merger history and the evolution of galaxies that live within them. Thus, in order to compare the clustering between SAM samples and HOD mocks which use only halo mass as input, it is necessary to remove

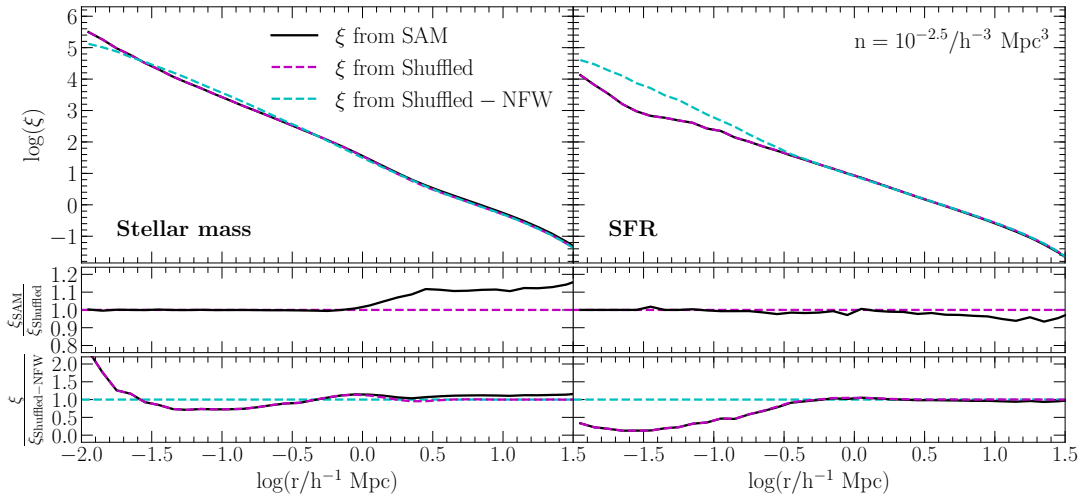

Figure 2.6: Top: Correlation functions of SAM samples (dotted black in the main panel, solid black in the subpanel), and their modifications: the shuffled (dashed magenta) and the shuffled-NFW samples (dashed cyan). The galaxy samples are selected by stellar mass (*left*) and SFR (*right*), with a number density of $10^{-2.5}h^3\text{Mpc}^{-3}$. Middle: Ratios between the 2PCF of the SAM and shuffled samples. Differences at large scales are signatures of assembly bias. Bottom: Ratios of the 2PCF with respect to the 2PCF from the shuffled-NFW samples. The differences in the one-halo term below 1 Mpc/h indicates the departure of the satellite profiles from a NFW.

the assembly bias signal in the former samples.

Assembly bias can be eliminated from SAM samples through the shuffling technique introduced by Croton et al. (2007). This consists of randomly exchanging the galaxy populations between haloes of the same mass, thus removing any connection to the assembly history of the haloes. This procedure does not change the distances from satellites to their central galaxy in each halo. In clustering terms, the one-halo term of this "shuffled" catalogue is the same as the original SAM sample but its two-halo term is different because assembly bias is not present in the shuffled sample. If the SAM samples did not have assembly bias, we would measure the same 2PCFs for their shuffled samples as measured for the original output.

Fig. 2.6 shows the correlation functions of a SAM sample and its shuffled version, for both the stellar mass and the SFR selected samples. The assembly bias signature, shown in the middle panels, is evident in the clustering differences between these two catalogues at large separations. We also show the 2PCF of a modified shuffled sample that will be introduced below. The assembly bias signatures remain unchanged for the other samples, but they are noisier for the lowest number density samples as they contain fewer galaxies.

It can be seen that assembly bias increases the clustering for stellar mass selected samples, as was shown by Zehavi et al. (2018). SFR selected samples, on the other hand, show a decreased clustering amplitude. For the intermediate galaxy density sample, the assembly bias enhances the two-halo term by $\sim$ 12% for the stellar mass selected sample and suppresses the amplitude in the SFR selection case by $\sim$ 4%. The enhance of clustering amplitude for the other stellar mass selections remains similar. For the SFR selections, we see that suppression on clustering becomes weaker for higher density samples. Indeed, assembly bias can enhance the amplitude if the density of the sample if very high, as shown in (Contreras et al., 2019).

2.3.5 The shuffled-NFW target catalogue: changing the satellite distribution in the SAM

The radial profile of satellites in the G13 SAM deviates from the standard NFW profile of dark matter within halos because the SAM associates galaxies with subhalos (or a proxy, such as the most bound particle, in the case of subhalos which are no longer resolved). The radial profile of subhalos is different from that of the dark matter (see e.g. Angulo et al. 2009). The choice of which subhalos (and former subhalos) are associated with galaxies is driven by the galaxy formation model, which determines the luminosity of any galaxy associated with a subhalo and whether or not it has merged due to dynamical friction (only type 2 satellites, those that no longer have a resolved subhalo associated with them, are considered as candidates for galaxy mergers).

The final step before testing the accuracy of the HOD models is to modify the shuffled SAM catalogue to force the satellites in each halo to follow an NFW profile. We call the result the shuffled-NFW catalogue. Because satellite galaxies in the SAMs and shuffled samples do not follow an NFW profile, the one-halo term of their 2PCFs are different to the one-halo term of the shuffled-NFW sample, as shown in the bottom panel of Fig. 2.6. The shuffled-NFW catalogue does not contain assembly bias, and the satellites follow the *same* NFW profile as adopted in the HOD mocks. We note that the shuffled-NFW is not intended to be the "best" prediction of galaxy clustering but rather is the target sample for the reconstructions using the HOD mocks which has a controlled 1-halo clustering pattern to facilitate testing.

2.4 Testing the accuracy of the HOD models

2.4.1 Satellite radial distributions and clustering of HOD mocks

In Fig. 2.7 we show the satellite profiles in the SAM samples, for stellar mass and SFR selected samples separated into the contributions from type-1 satellites and from orphan galaxies. To examine the departure from NFW, we produce a SAM-NFW catalogue where satellites in SAM are forced to follow the same NFW profile as used in the HOD mocks. For this catalogue, we update the satellite positions in the SAM samples according the same NFW density profile used to produce the HOD mocks. Note that this SAM-NFW is different from the shuffled and shuffled-NFW catalogues mentioned above.

It can be seen that the NFW profile is different from the profile of type-1 satellites and orphans, particularly for the SFR selected sample. The profiles in the 2-HOD and 4-HOD mock catalogues are also shown, and they match with the NFW as expected from the construction of the HOD mock. Both models also reproduce the NFW profile in the other number density samples.

The masses of host haloes of 1-HOD satellites do not correspond with the masses in the original SFR selected samples. Thus, the virial radii of these haloes define NFW density profiles that are different from the profiles in the other models. This has an impact on the positions of satellites generating the striking difference with NFW in Fig. 2.7 for the SFR selection. The same occurs for the stellar mass selections but it is less extreme than the SFR case.

The HOD models predict different galaxy populations for the G13 SAM samples. Table 2.2 shows the satellite fraction of the SAM samples and HOD mocks built by the three different models, assuming a Poisson distribution for the HOD of satellites. Note that the 2-HOD and 4-HOD predict almost the same satellite fraction as in the SAM samples because of the separate HOD modelling of centrals and satellites.

We check the accuracy of each HOD model by comparing the HOD mocks with the shuffled-NFW sample via their 2PCFs. Fig. 2.8 shows the clustering of the shuffled-NFW and the HOD mocks built using the HOD models described in Sec. 2.3.1. These particular models assume a Poisson distribution for the HOD of satellites. It can be seen that the three schemes produce accurate clustering predictions on large scales. On small scales, the 2-HOD and 4-HOD models produce similar accurate results while the 1-HOD shows striking differences. These deviations

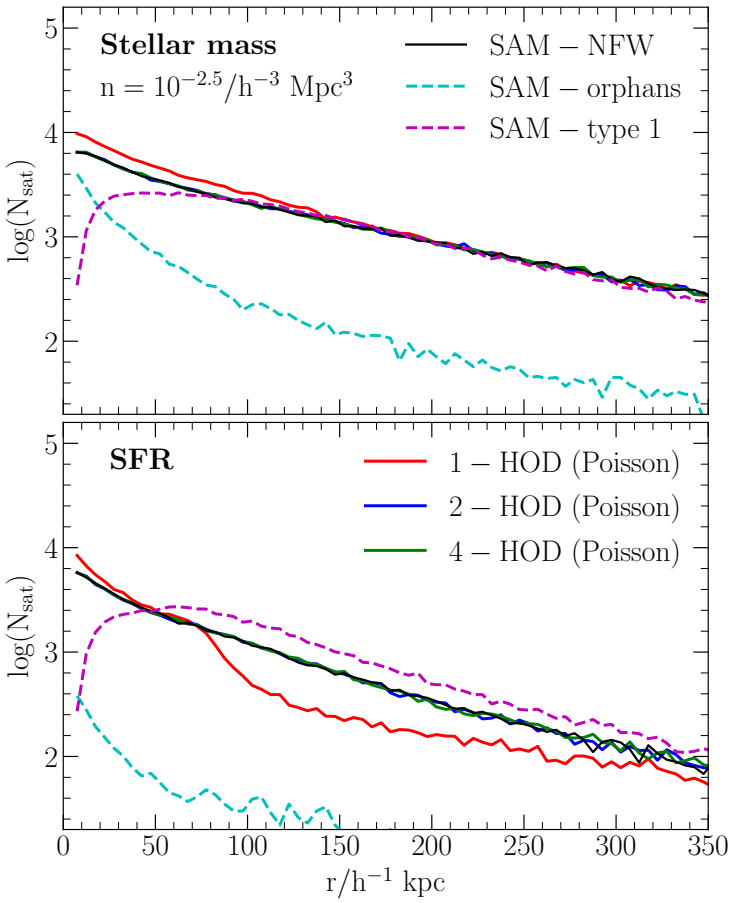

Figure 2.7: Profile of satellites hosted by subhaloes (dashed magenta) and orphans (dashed cyan) in a stellar mass (top) and SFR selected sample (bottom). The lines show the SAM with an NFW imposed for all satellites (solid black), and HOD mocks built by the 1-HOD (solid red), 2-HOD (solid blue) and 4-HOD models (solid green) where the HODs of satellites are described by the Poisson distribution.

Stellar Mass				
$n/h^3\mathrm{Mpc}^{-3}$	SAM	1-HOD	2-HOD	4-HOD
0.001	0.280	0.317	0.279	0.280
0.00316	0.322	0.405	0.324	0.324
0.01	0.381	0.517	0.383	0.383
SFR				
$n/h^3\mathrm{Mpc}^{-3}$	SAM	1-HOD	2-HOD	4-HOD
0.001	0.171	0.084	0.175	0.172
0.00316	0.195	0.192	0.197	0.197
0.01	0.244	0.334	0.246	0.246

Table 2.2: Satellite fractions of the galaxy samples used. The first column indicates their number densities. Column 2,3,4 and 5 show the satellite fraction in the SAM samples and in the HOD mock built using the 1-HOD, 2-HOD and 4-HOD models, respectively.

come from the overprediction of the number of satellites in the stellar mass selected samples. For the SFR selection cases, the difference is due to the notably different occupation function of central and satellites in the 1-HOD mock. The inaccuracy of the 1-HOD modelling is also present for the other number density samples too, whereas the 2-HOD and 4-HOD models produce similar quality results to the one shown here. As the 2-HOD and 4-HOD models are clearly the best, we drop the 1-HOD model henceforth.

2.4.2 Impact of the assumed HOD scatter

To study the impact of the scatter of the HOD on the clustering, we consider different dispersions for the negative binomial distribution in the construction of HOD mocks. Fig. 2.9 shows the 2PCF of HOD mocks using different β values. For the stellar mass selected samples, the scatter of the HOD does not have a significant impact.

For the SFR selection, the amplitude of the clustering on small scales is very sensitive to the scatter in the number of satellites. We find that increasing β changes the amplitude of the one-halo term. When we split the contribution to the clustering from low and high mass haloes, we observe that the scatter mainly impacts the one-halo term of low mass haloes. This feature is present in HOD mocks built using the 2-HOD and 4-HOD methods. This is reproduced by both HOD models, indicating that this is a feature particular to the SFR selected samples.

The most accurate clustering reconstructions for the G13 samples are obtained when we use the 2-HOD or 4-HOD to build mock catalogues assuming a negative binomial distribution for the HOD of satellites. Note that clustering predictions from both models do not show significant differences.

Fig 2.10 shows the particular results from the 4-HOD modelling for all the space density samples. For the G13 SFR selected samples, the 4-HOD (and the 2-HOD) modelling produces the best results when $\beta = 0.05$, which corresponds to a distribution slightly wider than a Poisson distribution. For the case of stellar mass selected samples, the best reproduction is obtained with $\beta = 0.08$. Using instead the Poisson distribution (i.e, $\beta = 0$) produces worse results for both selections particularly in the one halo regime.

For SFR selections, when using $\beta = 0.05$ and $\beta = 0$, the departures from the shuffled-NFW catalogues are below $\sim 8\%$ and $\sim 15\%$, respectively. It can be seen that the dispersion of the 2PCFs becomes important in the lowest number density sample. However, the assumption of the negative

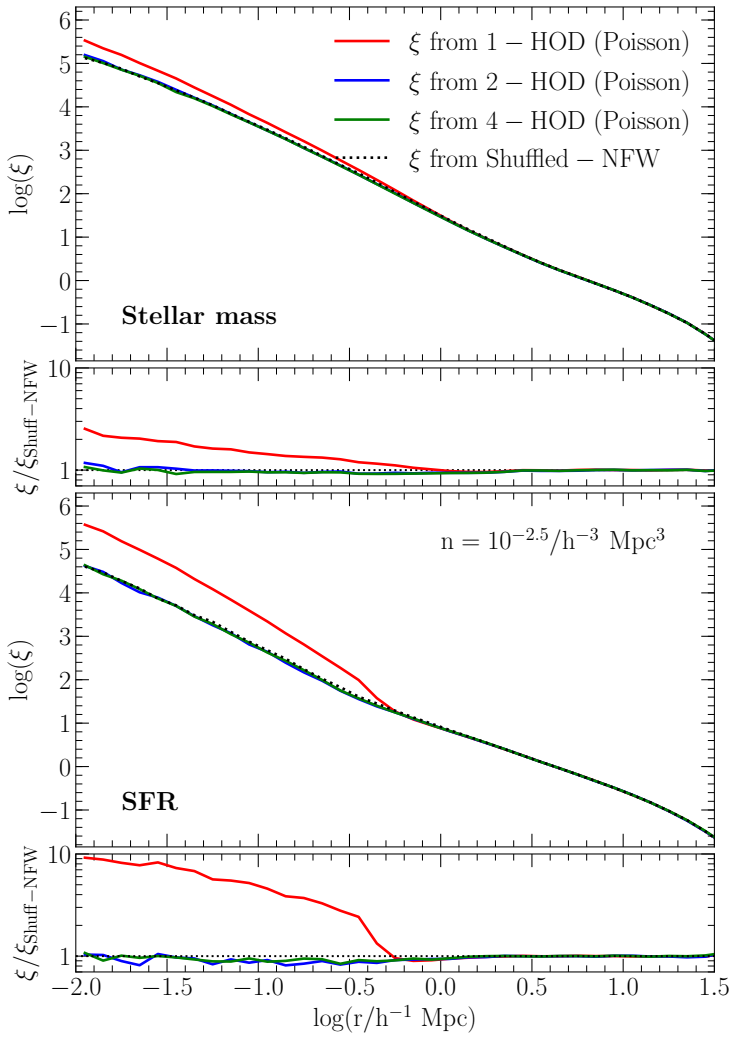

Figure 2.8: Clustering of HOD mock catalogues of stellar mass (top) and SFR selected samples (bottom) for a number density of n = $10^{-2.5}/h^{-3}$Mpc^3. The mocks are built using the 1-HOD (red), 2-HOD (blue) and 4-HOD (green) models, assuming a Poisson distribution for the HOD of satellites. The clustering of the shuffled catalogue with an NFW profile is shown as the dotted line. Subpanels show the ratios of the 2PCF of the mocks with respect to the 2PCF of the shuffle-NFW catalogue

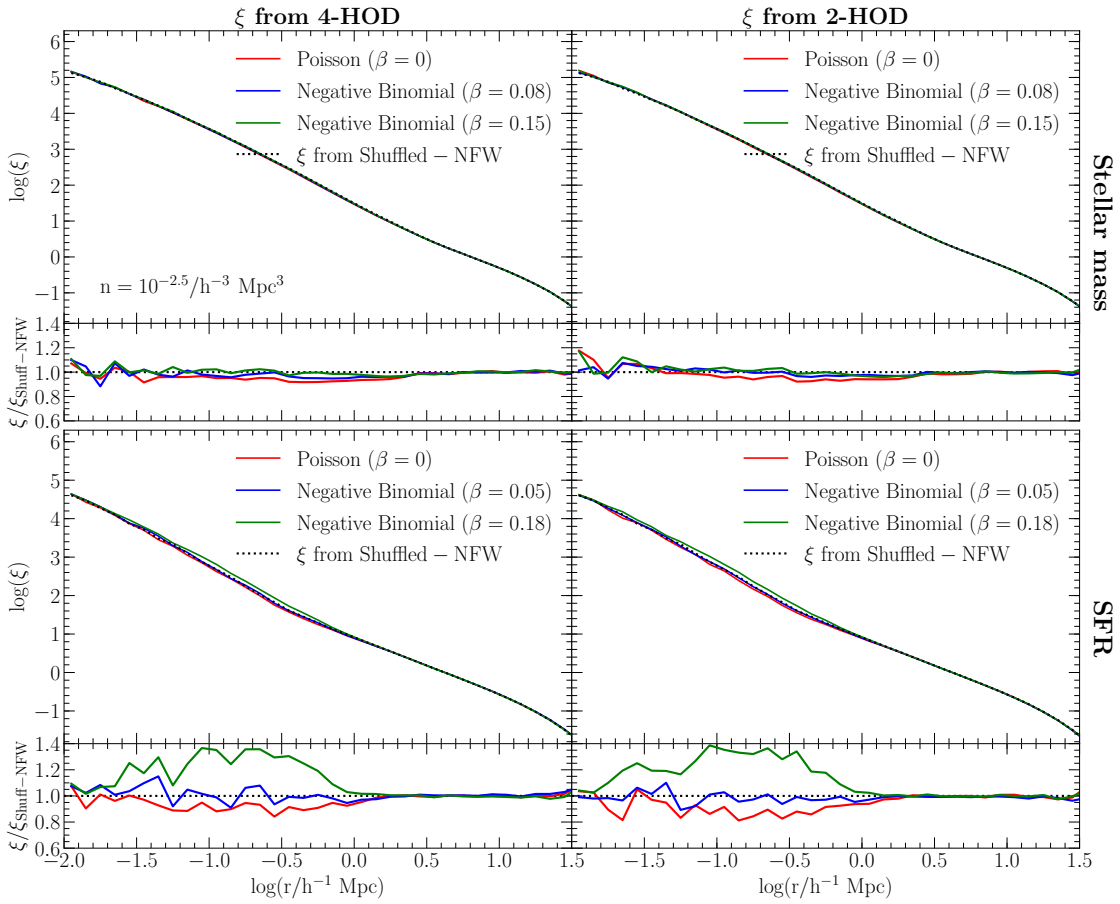

Figure 2.9: 2PCF of HOD mocks built using the 2-HOD (right) and 4-HOD (left) models, for stellar mass (top) and SFR selected samples (bottom). The HOD models are used to build mock catalogues assuming a Poisson distribution (solid red) and negative binomial distributions of $\beta = 0.05$ (solid blue) and $\beta = 0.18$ (solid green) for the HOD of satellites. The clustering of the shuffled-NFW samples (dotted black) is shown in each panel. The ratios between the clustering of HOF mocks and shuffled-NFW are shown in the subpanels.

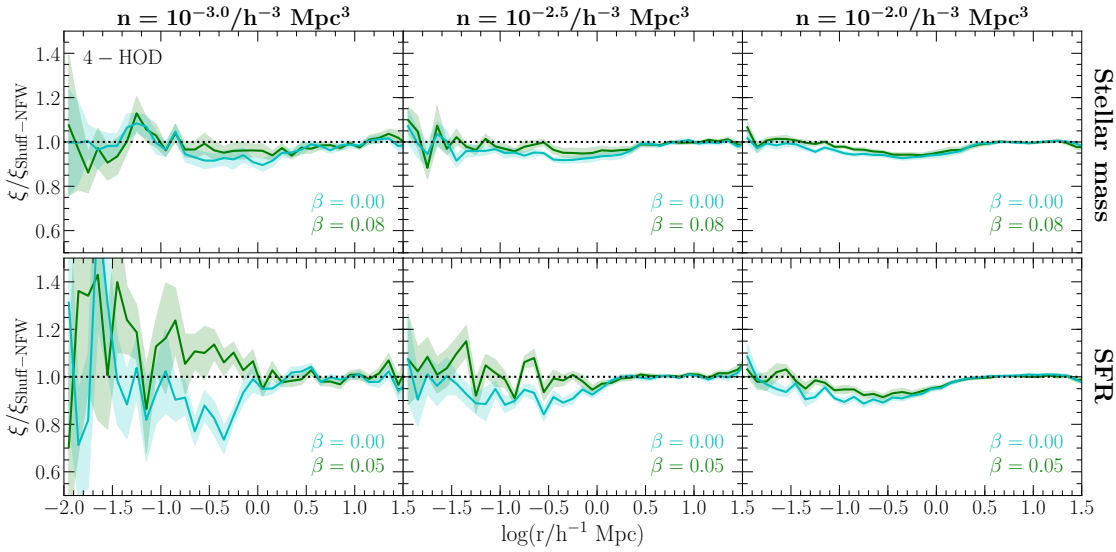

Figure 2.10: Ratios between the 2PCFs of mock catalogues, constructed with the 4-HOD method, and the shuffled-NFW catalogue. The HOD of satellites in the HOD mocks follows either the negative binomial (green) or Poisson distributions (cyan), with the color indicating the value of β. We show results for stellar mass (top) and SFR selected samples (bottom). Number densities increase from left to the right as labelled at the top of each column. The shaded regions represent jackknife errors calculated using 10 subsamples.

binomial distribution still produces better results, especially in the transition from the one- to the two-halo term.

For the stellar mass selection cases, the impact on clustering when using different β values is much less significant. Indeed, the weak relation between clustering and HOD scatter, shown in Fig. 2.9, suggests that it is not necessary to include additional scatter in the construction of HOD mock catalogues for stellar mass selections. To compare with the Poisson distribution, we show also the clustering prediction for the stellar mass samples using $\beta = 0.08$, for which we obtain the best result.

Satellites in the G13 SAM are well described by a non-Poisson distribution. This is consistent with the HOD of subhaloes found in Boylan-Kolchin et al. (2010). The recipes that build mock catalogues of SFR selected samples using the HOD approach must undertake an analysis of the scatter of the HOD of satellites as it impacts the clustering. This analysis will provide the best β to construct a HOD mock of a particular sample. For stellar mass samples, the HOD scatter has a weak impact on clustering, so the same analysis is not necessary in the context of HOD mock catalogues.

In this chapter, we have extended the HOD approach to building mock catalogues by including the scatter of the satellite's HOD as an additional parameter in the modelling. For galaxy samples from G13 semi-analytical model, the best mocks are produced using the HODs of central and satellites independently and assuming a negative binomial distribution for the HOD of satellites. The scatter has a significant impact on clustering, so it must be considered if HODs are used to study the future SFR or ELG selected samples.

Chapter 3

Clustering properties of ELG samples

In this chapter, we analyze the clustering of the emission line selected galaxies. The simulation and the definition of the galaxy samples are explained in Section 3.1. In Section 3.3 we show the galaxy scale-dependent assembly bias signature in [OIII] and [OII] line emission selected samples, and we compare them with the signals in SFR and Hα selections. In Section 3.4 we discuss a possible origin of this scale-dependence, and in Section 3.5 we explore the impact of this feature on the Baryon Acoustic Oscillation (BAO) and in a cosmological parameter. Note that, in this chapter, we use the galaxy samples (the SAG samples) from a different semi-analytical model (the SAG model).

3.1 Simulation data

To obtain clustering measurements with a suitable statistic at large scales, we use the semi-analytical model `SAG` (§ 3.1.1) applied in a cosmological volume 8 times larger than Millennium (§ 3.1.2). The code, used to obtain the emission lines, is analyzed in § 3.1.3.

3.1.1 The galaxy formation model: SAG

The semi-analytical code SAG follows the galaxy formation and evolution in the dark matter structures predicted by the N-body simulations. The first version of this code is presented in Cora (2006), and additional improvements have been applied (see Lagos et al., 2008; Tecce et al., 2010; Orsi et al., 2014; Muñoz Arancibia et al., 2015; Gargiulo et al., 2015). SAG takes into account

several physical processes as the radiative cooling of hot gas, the feedback effect from supernova explosions, the chemical enrichment of the gas, the growth of super-massive black holes that are related with AGN feedback effects, and the impacts in the gas components from galaxy mergers.

In previous versions of SAG, when a galaxy becomes a satellite all its hot gas is transferred to the hot gas component of the host halo, and in consequence star formation is suppressed. In the updated SAG code, presented in Cora et al. (2018), this scheme is modified to a gradual loss of the satellite's hot gas. Then, gas cooling is less affected, and the final state depends on the time-scale of this process. This is relevant for this work as more satellites can be more star-forming, and they can be included in an SFR-selected sample, or in an ELG sample.

The simulation provides the SFR and the instantaneous SFR for the bulge and disk components. The former value corresponds to an average of SFR between two snapshots. The instantaneous SFR is related to the values in the last sub-step of the simulation, just before the time epoch determined by the redshift. In the disk, the instantaneous SFR corresponds to a quiescent state, and in the bulge, it corresponds with starbursts episodes. Regarding the chemical enrichment, the SAG follows the production of elements from mass loss of massive stars or supernova explosions, which will be used for the next generation of stars. The SAG predicts the metallicity of the gas, which depends on the metals produced by such feedback effects.

Here we extract galaxies at $z \sim 1$ from the Cora et al. (2018) SAG version. We define our galaxy samples according to three different number densities where galaxies are ranked in a decreasing way (hereafter the SAG samples). The previous concepts on instantaneous SFR and metallicity are relevant because, as we will see, they are the inputs for the predictions of emission lines luminosities from the disk and bulge of galaxies.

3.1.2 The MultiDark Plank 2 simulation

The SAG is applied on the dark matter haloes extracted from the `MULTIDARK` simulation MDPL2 Klypin et al. (2016), which is publicly available in the `COSMOSIM` database[1]. MDPL2 uses the scheme of a ΛCDM universe, which is charaterized by the Plank cosmological parameters $\Omega_m = 0.307$, $\Omega_b = 0.048$, $\Omega_\Lambda = 0.693$, $h = 0.678$ and $n_s = 0.96$ (Planck Collaboration et al., 2014). The simulation follows 3840^3 particles within a cubic box of comoving side-length

[1] www.cosmosim.org

1 h^{-1} Gpc, with a mass resolution of $m_p = 1.51 \times 10^9 M_\odot\ h^{-1}$. The particles are followed from $z = 120$, where dark matter haloes are identified with the ROCKSTAR halo finder (Behroozi et al., 2013a), and the CONSISTENTTREES code (Behroozi et al., 2013b) is used to construct the merger trees, which are the starting point for SAG modelling.

The application of SAG into the haloes of the MDPL2 provides with a galaxy distribution in a very large cosmological volume. Thus, we can obtain clustering measurements with a good signal on very large scales, which is useful to study features in the large-scale structure of ELG samples.

3.1.3 The photoionization code and galaxy samples

The emission line luminosities from galaxies are not computed by SAG. We use the code implemented in Orsi et al. (2014) (hereafter O14) to obtain the Hα, [OIII]λ5007 and [OII]$\lambda\lambda$3727 − 3729 line emissions (hereafter Hα, OIII and OII, respectively) for all SAG galaxies at $z \sim 1$. The first and second emission lines correspond to visible radiation in the electromagnetic spectrum, while the [OII] doublet is the most energetic line, located in the near-ultraviolet regime. The neutral gas in the interstellar medium of galaxies absorbs a fraction of the radiation from young massive stars leading to the production of emission lines. Then, ELGs can be used as tracers of star-forming galaxies.

O14 model combines the photoionization code MAPPINGS-III (Dopita & Sutherland, 1995; Groves et al., 2004) and the SAG semi-analytical model to predict the nebular emission from a characterization of the ISM of galaxies. In particular, it relates the ionisation parameter of gas with the cold gas metallicity. This relation is used by MAPPINGS-III to compute the fluxes for a particular line. The Hα emission, in particular, is obtained from a linear relation with the instantaneous SFR.

We compute the emission from the bulge and disk components, and then we sum both quantities to obtain the total line emission luminosity of each galaxy. From this point, when we refer to line emission or line luminosity, we refer to the sum from both components. Hence, an ELG selection corresponds with selecting according to the total luminosity of a line.

We select galaxies in the SAG sample according to the emission luminosities of the three lines mentioned above. We also include stellar mass and SFR selections for future comparisons with ELGs. We define our galaxy samples according to the number densities $n = 0.001/h^3$ Mpc^3,

$n/h^3\mathrm{Mpc}^{-3}$	$\mathrm{M}^*_{\mathrm{min}}/h^{-1}\mathrm{M}_\odot$	$\mathrm{SFR}_{\mathrm{min}}/\mathrm{yr}^{-1}\mathrm{M}_\odot$
0.001	4.68×10^{10}	15.48
0.00316	1.58×10^{10}	7.05
0.01	0.29×10^{10}	1.68

Table 3.1: The number densities of the SAG samples are shown in the first column. The corresponding cuts in stellar mass and SFR are shown in the second and third columns, respectively.

$n/h^3\mathrm{Mpc}^{-3}$	$\log(\frac{\mathrm{L}_{\mathrm{H}\alpha}}{\mathrm{erg\ s^{-1}\ h^{-2}}})$	$\log(\frac{\mathrm{L}_{[\mathrm{OIII}]}}{\mathrm{erg\ s^{-1}\ h^{-2}}})$	$\log(\frac{\mathrm{L}_{[\mathrm{OII}]}}{\mathrm{erg\ s^{-1}\ h^{-2}}})$
0.001	42.44	42.15	41.83
0.00316	42.11	41.87	41.53
0.01	41.53	41.39	41.22

Table 3.2: Same to Table 3.1 but for Hα, OIII, and OII selections.

n = 0.00316/h^3 Mpc^3 and n = 0.01/h^3 Mpc^3. The resulting cuts for each number density are shown in Tables 3.1 and 3.2.

3.2 Characterization of the SAG galaxy samples

The 2PCFs of each galaxy sample are shown in Fig. 3.1. We see the same trends as for G13 where galaxies, selected by stellar mass, show larger clustering amplitudes than SFR and ELG selections. These differences are more significant in the 1-halo terms because the former sample contains more satellites. Hence, a larger number of galaxy pairs increase the clustering amplitude. For the other selections, we observe that selecting by more energetic lines results in less correlated samples at large scales. This indicates that the large scale structure is traced differently by the galaxies in each ELG sample (we will discuss this point in § 3.3).

Fig. 3.2 shows the halo occupation functions (HOD) of the galaxy samples for two number densities. As can be seen the HODs for the stellar mass selections follow the same canonical form for centrals and satellites. Moreover, we see that the shape of the HODs for emission line selections are similar to the SFR cases as expected because ELGs are tracers of star-forming galaxies (e.g. Gonzalez-Perez et al., 2018). Nevertheless, we found that ELGs tend to be in low mass haloes, which is consistent with O14 results.

Additional differences between the SAG samples can be found when we analyze how galaxies are located in the metallicity-instantaneous SFR parameter space of the disk component. This is

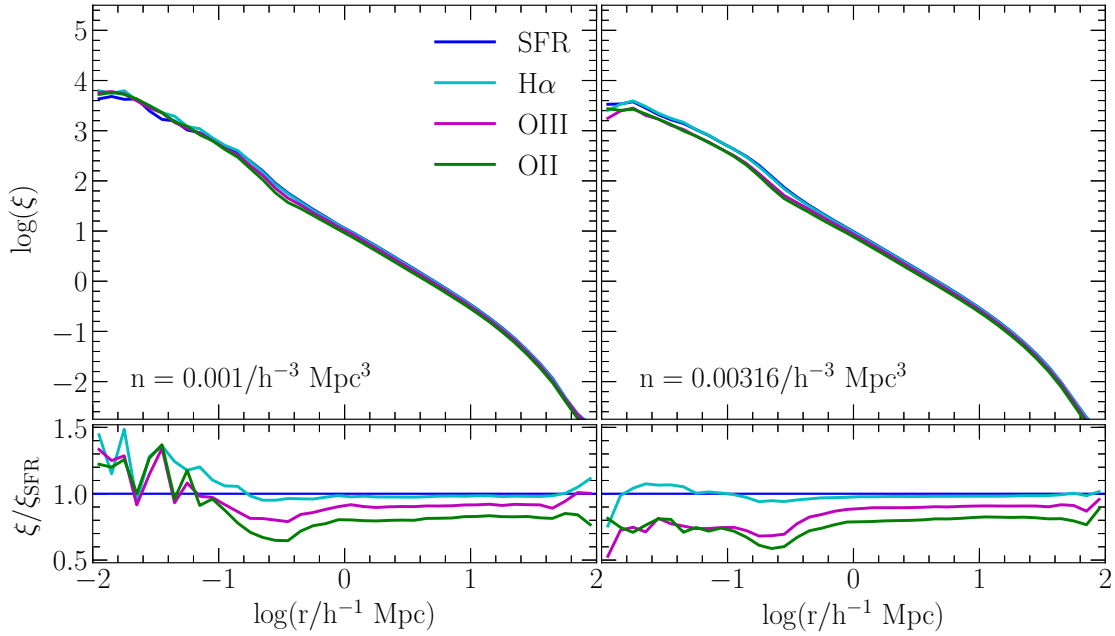

Figure 3.1: Similar to Fig. 2.1 but for the SAG samples with number densities n = 0.001/h^{-3} Mpc^3 (left) and n = 0.00316/h^{-3} Mpc^3 (right). The bottom panels show the ratio of clustering of the samples with respect to the SFR selected sample.

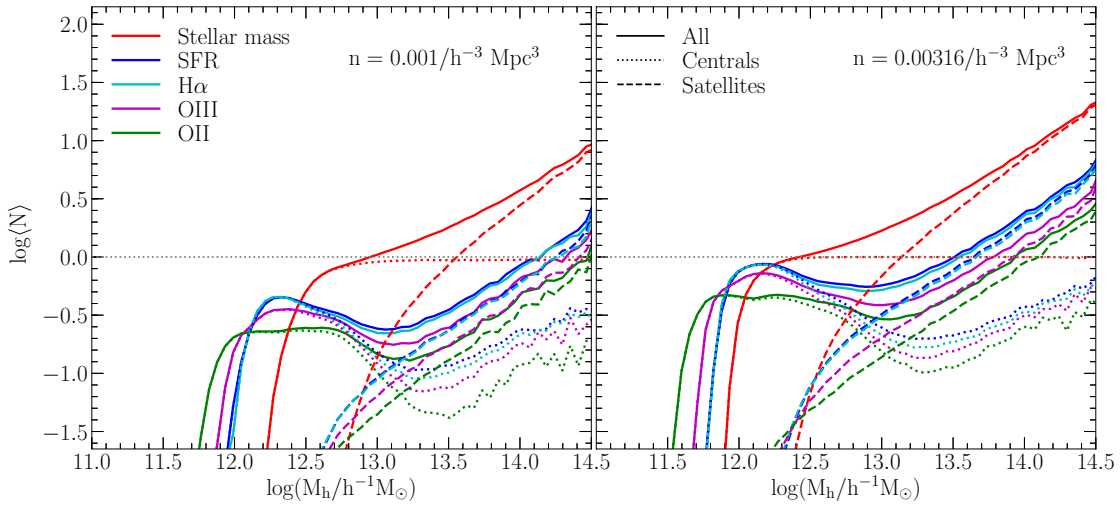

Figure 3.2: (Top) Similar to Fig. 2.2 but for the SAG samples with number densities n = 0.001/h^{-3} Mpc^3 (left) and n = 0.00316/h^{-3} Mpc^3 (right). We include the HODs for stellar mass selected samples (red lines).

shown in Fig. 3.3 for the stellar mass and the ELG selected samples. ELGs have similar distributions for instantaneous SFR showing similar positions for the mean and a small scatter. For the case of metallicity, we found that the OII selection contains more metal-poor galaxies than the other ELG selections. In the stellar mass selection case, we can see remarkable differences in the 2D distributions. These differences are expected as there is not a clear relation between stellar mass and line emission luminosities. Even though the last SAG version include an extended SF period for starburst in the bulge, the measured instantaneus SFR in zero for most galaxies. Then, we do not do the same analysis using the bulge parameters as there is no production of emission lines in most of the bulges.

3.3 The scale-dependent assembly bias

In simulations, the galaxy assembly bias (GAB) of galaxy samples selected by stellar mass, and more recently SFR, has been studied in SAMs (see Chapter 2) and Hydrodynamical simulations (Artale et al., 2018; Xu & Zheng, 2018). Even though ELGs can be related to star-forming galaxies, there are no studies about the GAB of ELGs. As the emission lines can be used to trace the SFR of a galaxy (either in simulations and observations), we expect that an important fraction of star-forming galaxies are also ELGs. Then, both spatial distributions are related.

One way to estimate SFR is through the detection of Hα emission from cold gas in the ISM. In SAG, we can use this relation in the opposite form, namely using the computed line emissions to study features in the large scale structure of ELG selected samples.

We apply the shuffling technique to the SAG samples to remove the dependence of clustering on secondary parameters. We determine the GAB signature from the ratio between the 2PCFs of the SAG samples and the shuffled samples. Because of the similarity between the ELG and SFR selections, we expect that GAB suppresses the clustering amplitude of the former. Fig. 3.4 shows the GAB signature of SFR, Halpha, OII and OIII selected samples, for two number densities. The signal for SFR and Hα selections are in close agreement, as expected from the close relation of their HODs (see Fig. 3.2), and because of Hα emission trace, in a good way, star formation in galaxies. For the OIII and OII cases, in contrast, we see significant deviations with respect to the other selections. In particular, their signals contain a bump in the transition from the 1- to the 2-

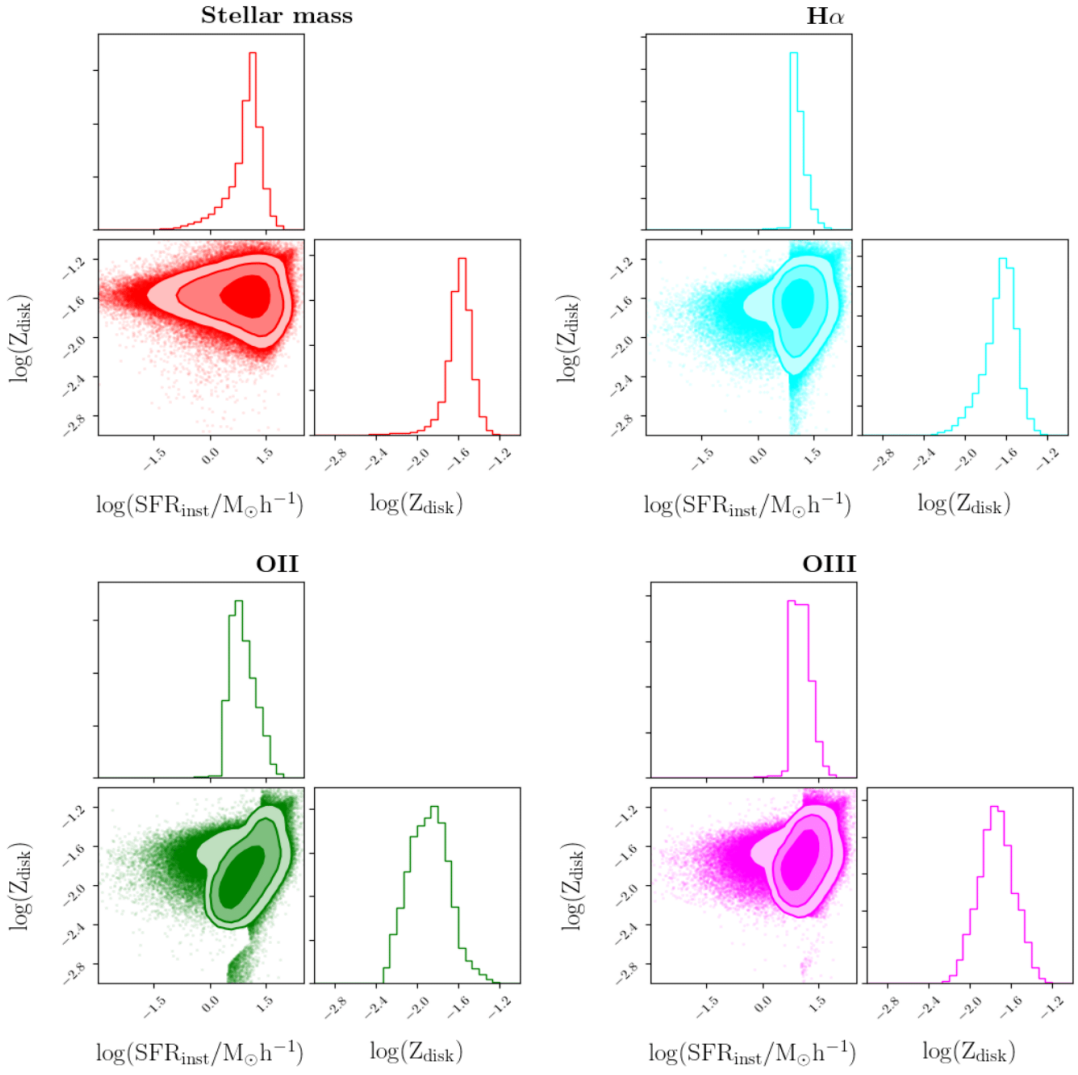

Figure 3.3: The figure shows the location of SAG galaxies in the metallicity-instantaneous SFR parameter space. The galaxy samples are selected by stellar mass (upper left), Hα (upper right), OIII (lower left) and OII (lower right), with a number density of n = 0.00316/h^{-3} Mpc^3. The different color levels represent 68%, 95% and 99% of the 2D distributions. The single distribution for Z_{disk} and instantaneous SFR are included for each selection.

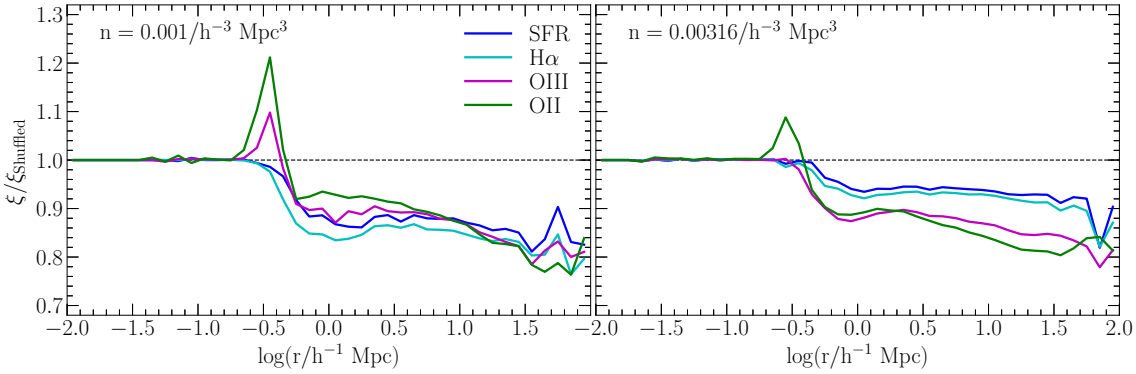

Figure 3.4: The galaxy assembly bias signature for the SAG samples with number densities (left) n = $0.001/h^{-3}$ Mpc^3 and (right) n = $0.001/h^{-3}$ Mpc^3. The OII and OIII selections show a scale-dependent signal for both number density samples.

halo term regime. Furthermore, the signature is scale-dependent for r $\gtrsim$ 1 Mpc/h, and it is steeper for the lowest number density (i.e for the galaxies with the highest line-emission luminosities).

The scale-dependence is a feature in the large-scale structure of ELGs of the SAG model. We check if this is present in G13 galaxies using the O14 photoionization code to obtain the nebular emission, but considering the (non-instantaneous) star-formation rate and gas metallicity predicted by G13. Fig. 3.5 shows the resulting GAB signature for OII and OIII selected samples. We show measurements for log(r/Mpc h^{-1}) < 1.5 as the cosmological volume is 8 times smaller than the MDPL2. The scale-dependent signature is reproduced at large scales, which indicates that this effect is a property in the large-scale structure rather than a particular result from a SAM modelling. Moreover, using SFR as the input parameter of the O14 model, also reproduce the scale-dependence making metallicity the main component tracing this effect.

To separate the contribution to the scale-dependent signal, we compare the 2PCFs of the SAG samples and shuffled samples with the clustering of dark matter. The latter is obtained applying the cosmological parameters of the MDPL2 simulation to the `CAMB` code[2], which compute the amplitude of density fluctuations as a function of scale. This calculation is purely cosmological and does not depend on any baryon physics, so it is independent of the predictions of the galaxy formation model. We use the clustering of dark matter to compute the bias parameter of the galaxy samples using equation (3.1). The value of this parameter is different for different selections and quantifies how galaxies trace the underlying dark matter density field. In a first order prediction,

[2]https:lambda.gsfc.nasa.gov

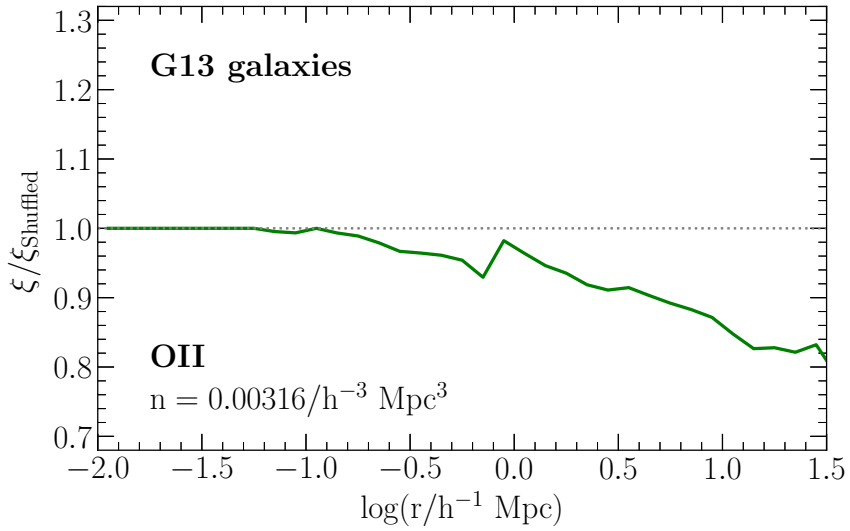

Figure 3.5: Same as Fig. 3.4, but for G13 galaxies at $z = 0$ with n $= 0.00316/\mathrm{h}^{-3}$ Mpc^3, and for a OII selection only. The OII emission was computed using SFR instead of instantaneous SFR. The scale-dependent is reproduced indicating that its origin should be related to a special feature on the large-scale structure of OII selected galaxies rather than a specific SAM modelling.

where the evolution of a perturbation is linear, the bias is constant at large scales.

$$b = \sqrt{\frac{\xi_{\rm gal}}{\xi_{\rm DM}}} \tag{3.1}$$

Fig. 3.6 shows the 2PCFs of the dark matter, the SAG sample, and the shuffled sample, for different selections. For clarity, we show results for the 2-halo regime only ($\log(\mathrm{r/Mpc\ h^{-1}}) \gtrsim 0$). The color lines in the bottom panels show, the bias parameters computed using equation (3.1). We averaged these parameters between 25 and 50 Mpc h^{-1} to obtain a constant large-scale bias that we compare with $b(r)$. We found that the bias, for the SAG and shuffled samples, are roughly constant in $25 < \mathrm{r/Mpc\ h^{-1}} < 50$ for the Hα selection. For the OII and OIII selected samples, the bias parameters have a significant dependence with the scale-separation, indeed the bias for the OII selection shows a strong steep relation. The bias, for the shuffled samples, also show a scale-dependence, but they are less significant than the unshuffled versions. The larger values for the bias found in the Hα selections indicates that these ELGs trace higher peaks in the density field than OIII or OII selected galaxies.

To figure out which galaxies contribute more to the scale-dependence, we split galaxies accord-

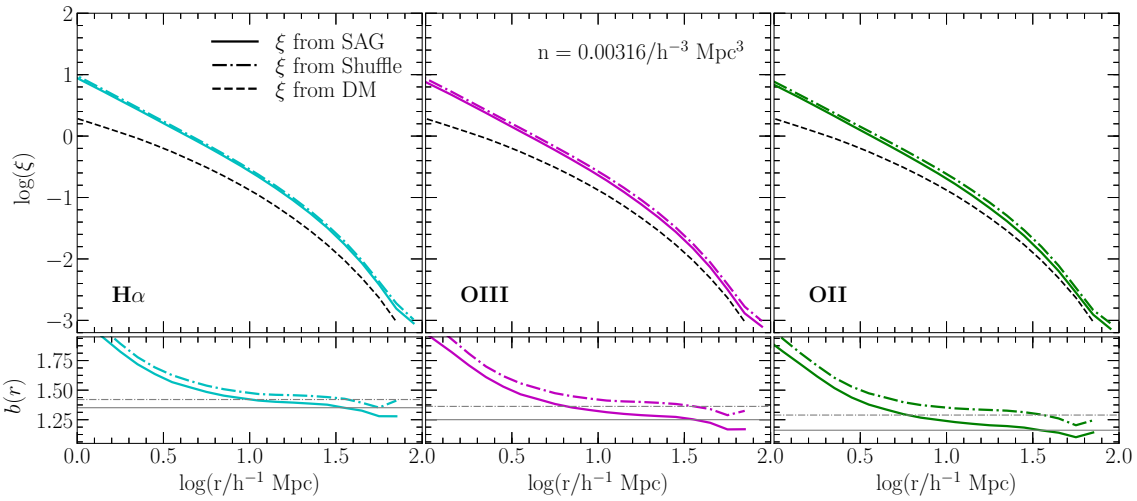

Figure 3.6: (Top) The 2PCFs of the SAG samples (solid) and the shuffled samples (dashed-dotted) for the Hα (left), OIII (middle) and OII (right) selections. (Bottom) The bias parameter, as a function of scale-separation, for the SAG and shuffled samples. The horizontal solid (dashed) grey line corresponds to the bias parameter of the SAG (shuffled) sample averaged between 25 and 50 Mpc h^{-1}.

ing to the mass of their host haloes. Fig. 3.7 shows the auto-correlation functions of OII and OIII sub-samples, where galaxies are hosted by low and high mass haloes. The scale-dependent GAB signature is reproduced for galaxies in low-mass haloes.

3.4 A possible origin of the scale-dependent GAB

3.4.1 Large-scale properties from gas metallicity

Here we look into possible origins of the scale-dependence GAB signature, found for OII and OIII selected samples. Fig. 3.8 shows the OII and Hα line emission luminosity as a function of the input parameters used in O14 model, namely instantaneous SFR and gas metallicity. For comparison, we also show the dependence with halo mass. In order to obtain a consistent signal of the median of the distributions, we do not include the galaxies with zero metallicity. The Hα emission indicates a one-to-one relation with instantaneous SFR, as expected because of the connection between SFR and Hα emission. In contrast, the OII emission exhibit a strong dependence on both SFR and metallicity. The latter plays a significant role in the prediction of OII (and OIII) emission. Then, we expect that the OII and OIII selected samples contain features from this metallicity dependence that may potentially affect the galaxy clustering.

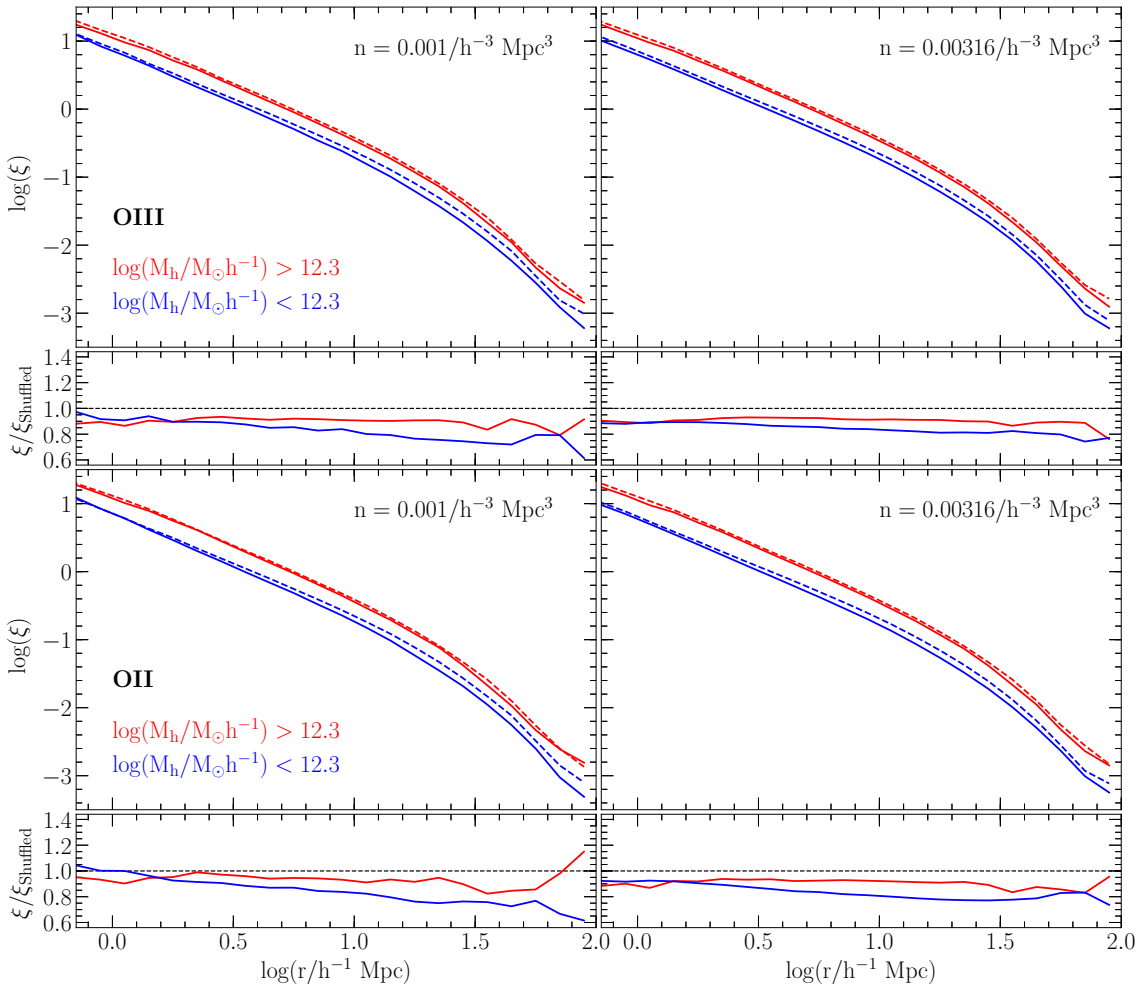

Figure 3.7: The main panels show the auto-correlation functions of OIII (top) and OII (bottom) SAG subsamples (solid) and their shuffled version (dahsed). The subsamples contain galaxies hosted by haloes with $\log(M_h/M_\odot\ h^{-1}) < 12.3$ (blue) and $\log(M_h/M_\odot\ h^{-1}) > 12.3$ (red). The original SAG samples correspond to number densities of n = 0.001 h^{-3} Mpc^3 (left) and n = 0.00316 h^{-3} Mpc^3 (right).

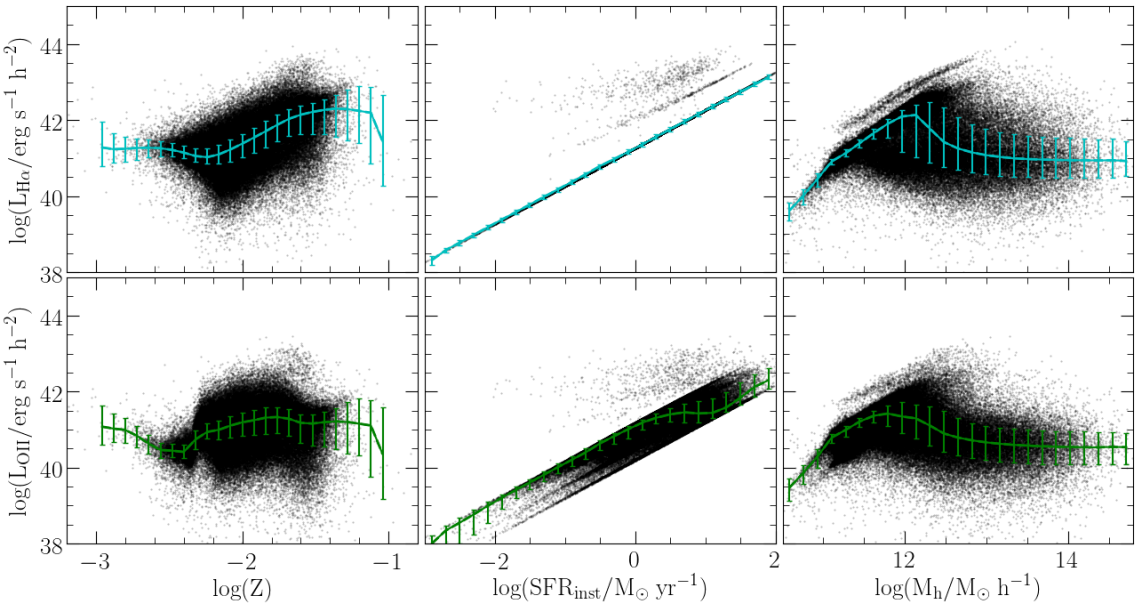

Figure 3.8: The Hα (top panels) and OII (bottom panels) line emission as a function of metallicity (left), instantaneous SFR (middle) and halo mass (right). The error bars represent the 20th and 80th percentile of the distribution in each bin.

n/h^3Mpc^{-3}	Z_{max}
0.001	0.0038
0.00316	0.0051
0.01	0.0073

Table 3.3: The first column indicates the number density of the samples. The resulting metallicity cuts are shown in the second column.

To determine if metallicity is tracing the scale-dependence, we produce SAG samples where galaxies are ranked according to their metallicity in a *increasing* way. Table 3.3, shows the resulting metallicity cuts for samples with different number densities. This definition has the advantage that the lowest density sample contains the most metal-poor galaxies which can contribute to the scale-dependence as they are hosted by low mass haloes (see. Fig. 3.7) that produce the steep signal. We construct the shuffled versions of each sample to determine their GAB signature. Fig. 3.9 shows the GAB for these metallicity selected samples. It can be seen that the scale-dependence is recovered, moreover when selecting galaxies with low and intermediate metallicities (left and middle panels) a steep signal is observed. This indicates that metallicity is tracing some particular feature in the large-scale structure of OII and OIII selected samples.

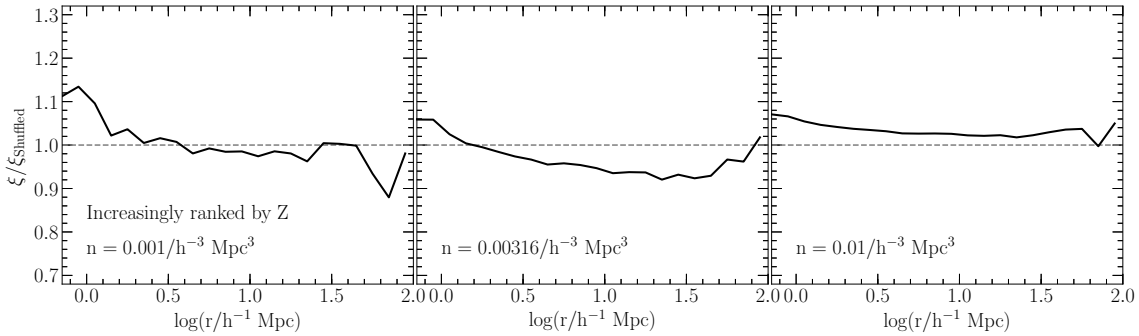

Figure 3.9: The GAB signature for metallicity selected samples of three number densities. In this case, the lowest (highest) density sample contains the most metal-poor (rich) galaxies. The scale-dependent assembly bias is reproduced for the two lowest density samples.

The metallicity distribution of the SAG samples are shown in Fig. 3.10. It can be seen that the mean of the distribution of OII and OIII selected samples indicates that these galaxies tend to be more metal-poor than the other selections.

In this section, we described some features in the OII and OIII selected samples that could be related to their strange GAB signature. In particular, galaxies with low and intermediate metallicities, and hosted by low mass haloes, are contributing to this non-typical GAB signature. Below, we discuss one possible explanation for this effect.

3.4.2 A selection by local density?

Previous works have found that galaxies in underdense regions have lower gas metallicities than galaxies in denser regions (Wegner & Grogin, 2008; Kreckel et al., 2015). When selecting by emission line luminosity, we may select by an environmental component as metallicity is related with features in the large scale structure. In particular, we expect that ELGs can be related to external properties of their host haloes. This additional environment component can be contributing to the scale-dependent GAB signature. For the stellar mass selections, on the other hand, haloes are selected by their halo mass instead of an environmental component.

To determine the GAB in each sample, we shuffle the galaxy population in haloes among haloes of the same mass to remove the impact on clustering from secondary properties. The shuffling technique takes into account only the mass to create the shuffled catalogues, so it is possible to change the original environment of galaxies to a more homogeneous distribution. Hence, if SAG

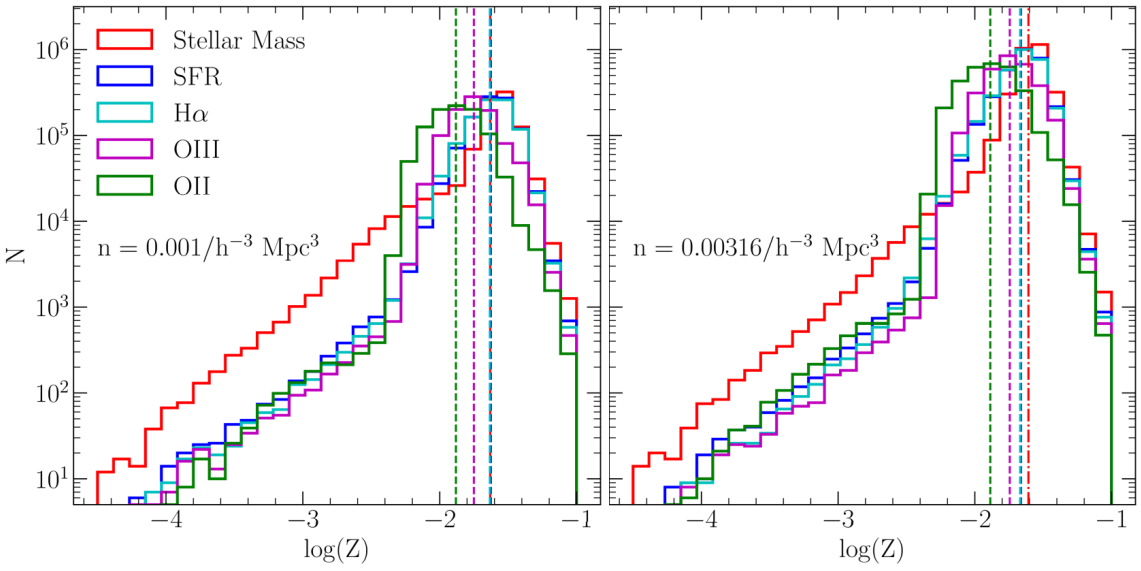

Figure 3.10: Metallicity distribution for stellar mass (red), SFR (blue), Hα (cyan), OIII (magenta) and OII (green) selected samples with two different number densities. The dashed vertical lines indicate the mean of each distribution.

galaxies, selected by a particular property, are mostly in underdense regions, they may be moved to haloes with the same mass but in more overdense environments. Moreover, the magnitude of this modification in the large scale structure may depend in the halo mass.

One approach to obtain measures of halo environments is computing the local number densities of the haloes that host the selected galaxies. To do so, we compare the halo fields in the SAG samples and in the MDPL2 simulation. We consider haloes with $M_h > 10^{10.8} M_\odot\ h^{-1}$ range[3]. Note that, in this way, we can compare the haloes in each selected sample with respect to the same field. The local number density for a halo is defined using its distance to the 5th closest halo (r_{5th}), then $n = 5/V(r_{5h})$ where $V(r_{5h})$. is the volume of a sphere with radius $r = r_{5th}$. Note that we only use host haloes and no subhaloes.

In the middle panels of Fig. 3.11, we show the average local number density for haloes in mass bins, for the original SAG samples and their shuffled versions. This plot must be analyzed in combination with the effective bias plot (what haloes contribute more to the clustering amplitude) in the upper panels, which gives us a clearer view of galaxy clustering taking into account the HODs, the halo mass function and the bias parameter. All of these parameters depend on halo

[3]This to avoid resolution problems that can produce very low mass haloes

mass, so the effective bias is different for each SAG sample[4]. This indicates, for example, that haloes, for the OII selected samples, with $M_h \sim 10^{12}$ $M_\odot$ h^{-1} contribute more to the clustering amplitude, and that we should only consider stellar mass results for $M_h \gtrsim 10^{12.3}$ for the lowest number density. The densities for the shuffled versions are similar for a wide range of halo masses, as expected because the shuffling tends to put galaxies in haloes of random environments. For the original SAG samples, there are clear differences, especially for the lowest density sample. The bottom panels, show the ratio between densities in the SAG samples and shuffled samples, as a function of halo mass. We see that $n_{SAM}/n_{Shuffled} \neq 1$ (and is scale-dependent), for a wide range of halo masses, indicating that the shuffling impacts the original environment of haloes in the OII sample. This is even more clear in the larger density samples where the ratio is scale-dependent for $11.7 < \log(M_h/M_\odot\ h^{-1}) < 12.1$. The same trends are observed for OIII selection but the dependence is weaker. The low number of low mass haloes produces the noisy signals of the average local densities for the stellar mass and SFR selections.

The local number densities in Fig. 3.11 represent a continuous set of values, as we are using different r_{5h} for each halo in the SAG samples. we seed differences in the local densities between the SAG and shuffled samples, which correspond with a change of environment. To distinguish whether this change is in the small or large scales regime, we re-compute the densities in spheres of fixed radius. For each halo, we obtain two local densities, which are defined by the number of objects in spheres of radii $r = 1$ Mpc h^{-1} and $r = 10$ Mpc h^{-1}. The former can trace features in the close external region of haloes with virial radii smaller than 1 Mpc h^{-1}, which correspond to haloes with $M_h \lesssim 10^{13}$ $M_\odot$ h^{-1}. The larger radius, on the other hand, can trace features of the halo distribution in larger scales, obtaining a clearer view of the environment of haloes.

The average local densities, as a function of halo mass, using these last two approaches, are shown in Fig. 3.12 for the SAG samples only. The bottom panels show the ratio between the measures in the middle panel and results for the SFR selection. For the analysis of this panel, we consider halo masses of $M_h > 10^{12}$ $M_\odot$ h^{-1}, where the effective bias for the SFR sample is significant in the lowest density sample. Using the large sphere case, we find that the local densities are roughly the same for most of the halo masses. Using the small sphere, we find that local densities depend on the specific SAG sample. The constant density for $M_h \gtrsim 10^{13.9} M_\odot\ h^{-1}$ is

[4] And the effective bias for each sample does not change after the shuffling as the HODs are preserved

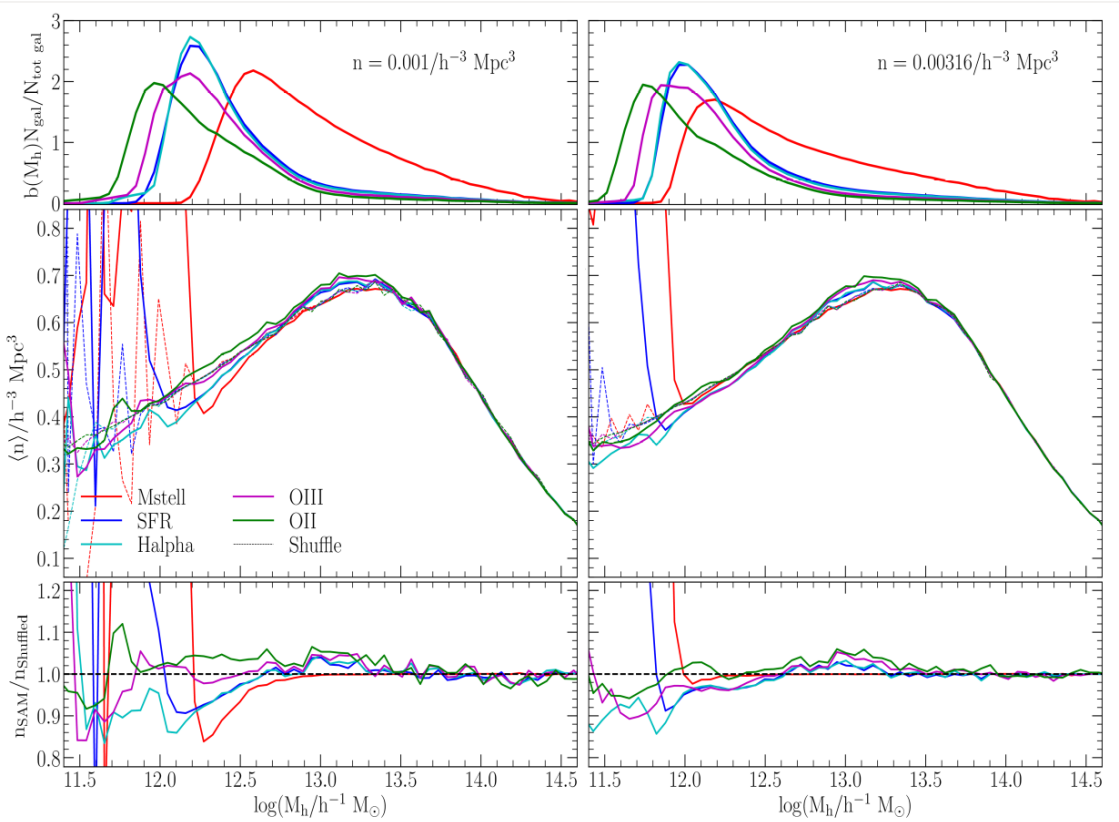

Figure 3.11: (Top) The contribution to the effective clustering bias as a function of halo mass. (Middle) the average local number densities for the SAG samples (solid) and their shuffled samples (dotted) as a function of halo mass. (Bottom) The ratio between the two measurements (SAG, shuffled) shown the middle panel. Different colours indicate different selections. Results for the lowest (left) and intermediate (right) density sample are shown.

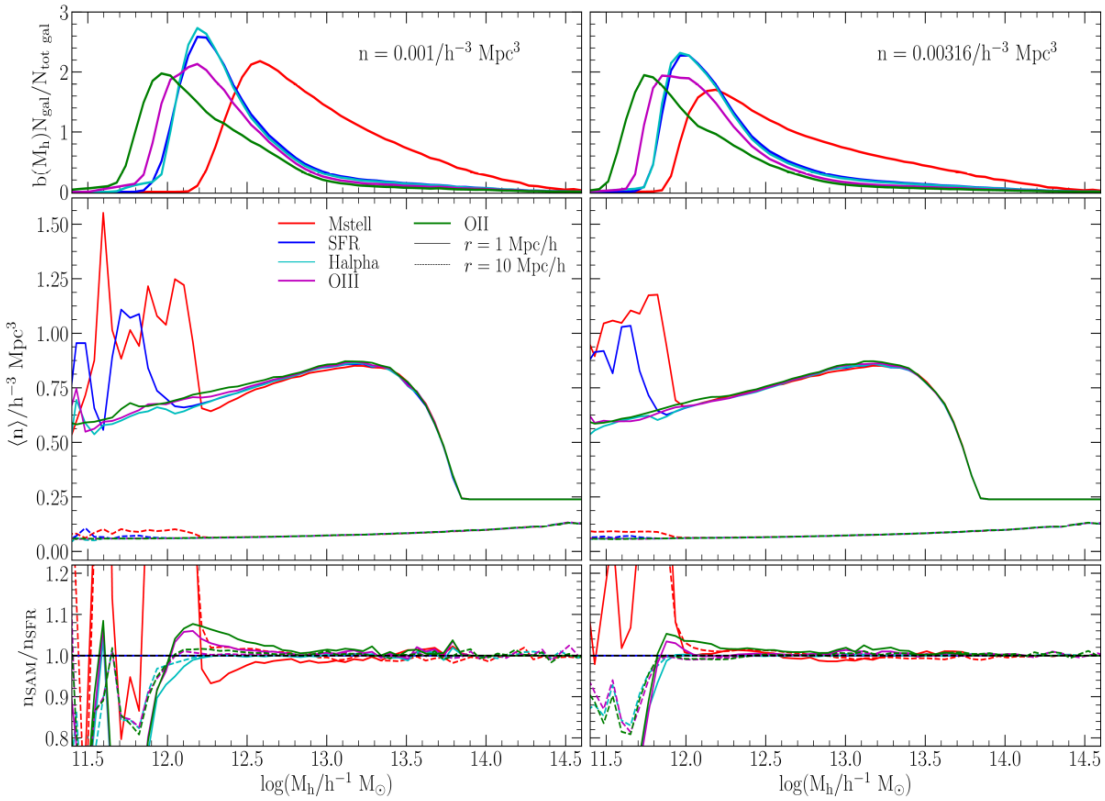

Figure 3.12: Same as Fig. 3.11 but considering SAG samples only, with local densities computed using the number of haloes in spheres of radii r = 1 Mpc h^{-1} (solid) and r = 10 Mpc h^{-1} (dashed). The bottom panels show the ratios between the local densities of haloes (using both spheres) in the SAG samples and results for SFR.

due to that the virial radii of those haloes are larger than r = 1 Mpc h^{-1} so the sphere only contains one object. For the r = 1 Mpc h^{-1} case, we see that the OII and OIII selected galaxies to show a scale-dependent relation with respect to SFR densities. Note that when comparing with Hα we obtain almost the same local densities. For the large radius case, we found smaller deviations from SFR densities. This indicates that there are special features in the close environment of the OII and OIII haloes. From this point, if we mention the local densities, we will be referring to the ones computed using the distances to the 5th closest halo.

The stellar mass, SFR and Halpha selected samples do not present a clear scale-dependent GAB signature (see Fig. 3.4). One possible explanation is that these samples do not contain (or contain in a low percent) the galaxies that contribute to the steep GAB signal. Fig. 3.13 shows scatter plots, color-coded by metallicity, of stellar mass-OII emission and Halpha-OII emission of galaxies from

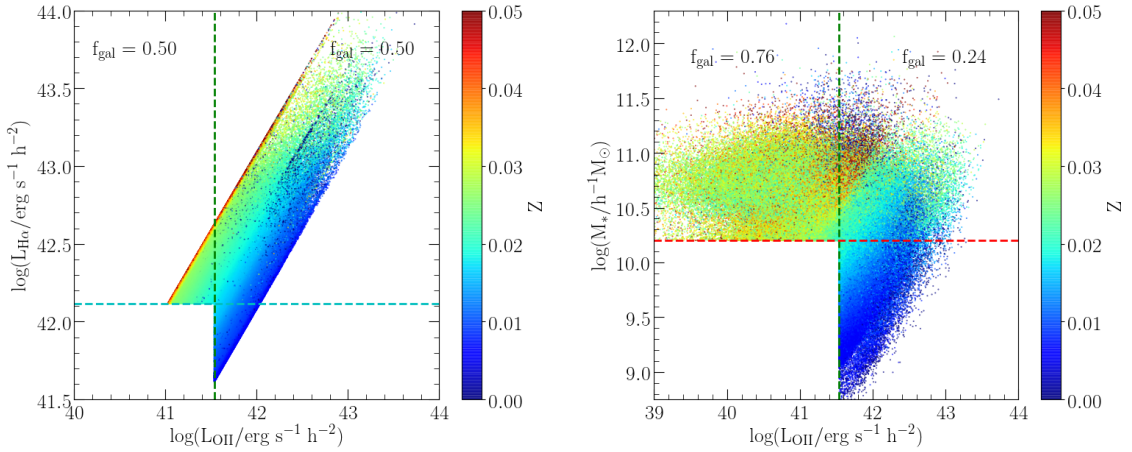

Figure 3.13: (Left) The Hα emission as a function of OII emission color-coded by metallicity. The cyan (green) dashed line indicates the cut in Hα (OII) for a sample of number density n = 0.00316/h^{-3} Mpc^3. The fraction of galaxies with OII emissions below and above the OII cut are included in both sectors. (Right) Same as the left but for stellar mass as a function of OII emission.

the original SAG simulation. The property cuts, shown as the vertical and horizontal dashed lines, correspond to the samples with a number density of n = 0.00316/h^{-3} Mpc^3 (see Tables 3.1 and 3.2). These lines define four different sectors in the property-space, where we can identify the galaxies selected by both properties or by just one of them. For example, for the stellar mass-OII plot, the stellar mass selected galaxies are above the red dashed line, and galaxies selected by OII and stellar mass at the same time are located in the upper-right sector.

There are more galaxies with low metallicity in the OII selection than in the other. This is consistent with the metallicity distributions shown in Fig. 3.10. As expected, there are overlaps between galaxies selected by each property. The half-part of galaxies selected by OII are contained in the Halpha selection. In contrast, a small fraction (24%) of the former sample is part of the stellar mass selection.

We compare the GAB signatures and the local number densities of the galaxies in each quadrant of the scatter-plots. Fig. 3.14 shows the same stellar-mass-OII and Halpha-OII relations of Fig. 3.13, but they are color-coded by its location in the figure. To obtain the GAB we compute the auto-correlation functions of each subsample, then we shuffle their galaxies producing the shuffled catalogues and, finally, we divide both quantities.

The GAB signatures, between each subsample (with the same color-coded), are clearly differ-

ent. In particular, for the lowest metallicity galaxies in the lower-right quadrant, we see a strong scale-dependent GAB. Even more, we found that significant fractions of the stellar mass and Hα selections (the blue-coded galaxies) contain this effect. Moreover, the local density distributions, for the grey-coded galaxies, show a larger population of galaxies living in low-density regions. This galaxy sample has the steepest GAB signature and contains the most metal-poor galaxies (see the lower right quadrant in Fig. 3.13).

Finally, we show the metallicity-local density relations for each sample in Fig. 3.13. we see that metallicity has a weak dependence on local density. Even though the dispersion is large, we see that emission line galaxies have lower metallicities, which is consistent with the previous results.

3.5 The impact on cosmology

In the previous sections, we have described the scale-dependent GAB signature in OII and OIII selected samples, and we investigate a possible origin of this effect. We found that this signal is, mainly, a contribution from galaxies with low metallicities that tend to live in underdense regions (see Fig. 3.13 and bottom panels of Fig. 3.14).

In this section, we look for impacts on cosmology from this scale-dependent signal. This is important as ELG data will be used by future surveys to constrain cosmology. Hence, any systematic that make a significant impact on the results from data analysis should be considered.

The Baryon Acoustic Oscillation (BAO) is a feature imprinted in the large scale structure of galaxies and is produced by the decoupling between baryons and photons during the recombination epoch. The scale of the BAO can be measured from galaxy surveys and can be used to constrain the nature of dark energy and cosmology (e.g Alam et al., 2017; Abbott et al., 2019). The BAO signals for the SAG samples and the shuffled versions are shown in Fig. 3.16 for two number densities. We multiplied the 2PCFs by r^2 to obtain a clearer view of the clustering amplitude. We find that the scale of the BAO peak is at lower scales for the OII selections than in the stellar mass case. This indicates that the large scale structure of these ELGs has special features that can be related to the scale-dependent signal discussed in the previous sections.

Furthermore, the peak of the OII shuffled version, for the highest number density sample, is clearly at larger scales than the position in the original sample. If this scale shift is real, we can

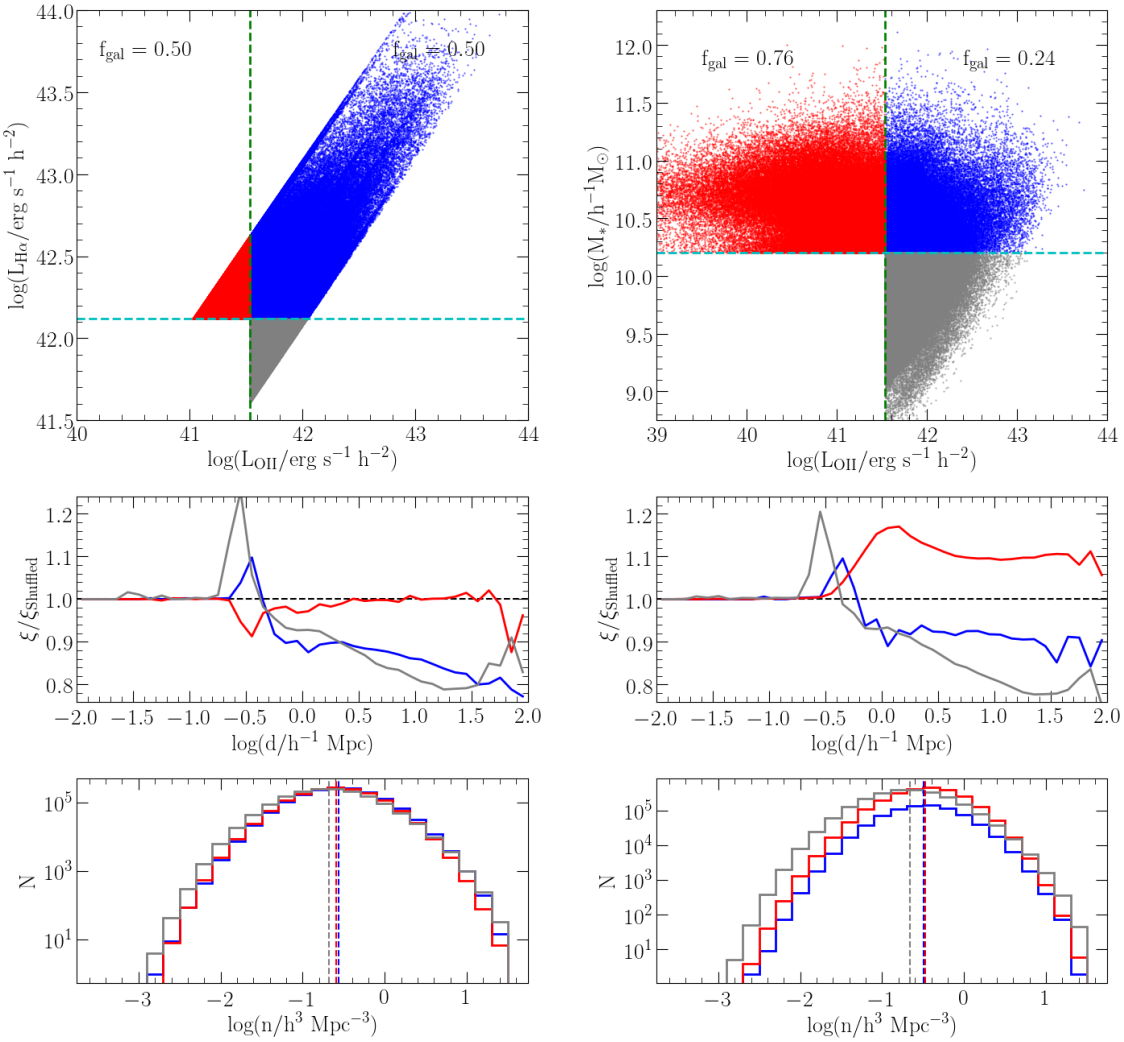

Figure 3.14: (Top) Same as Fig. 3.13 but with galaxies color-coded by their location in the main panel. (Middle) the GAB signatures for the three color-coded subsamples. (Bottom) The local density distribution for galaxies in each region in the top panel, where dashed vertical lines correspond to the mean of each distribution.

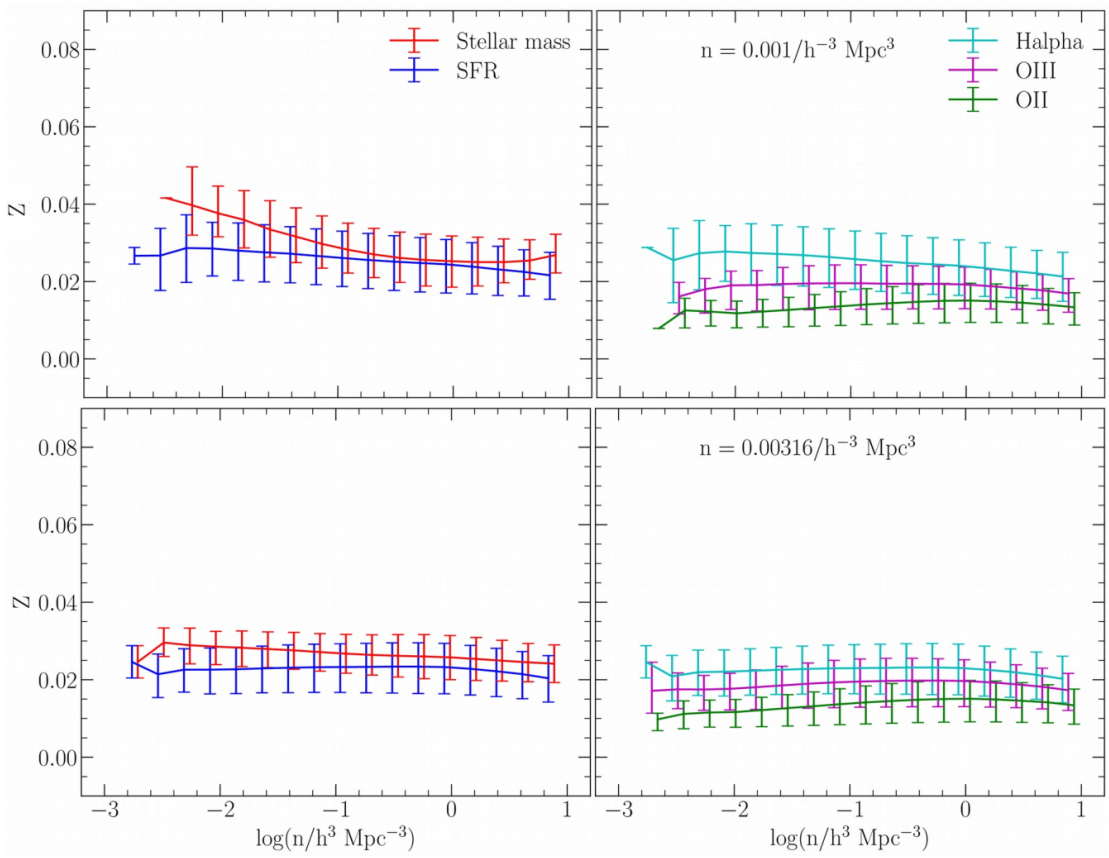

Figure 3.15: The metallicity of galaxies as a function of local density for SAG samples of number denisity n = 0.001/h^{-3} Mpc^3 (top) and n = 0.00316/h^{-3} Mpc^3 (bottom). The solid lines correspond to the median metallicity, and the bars indicate the 20th and 80 per cent range of the distribution. Note that ELGs tend to have lower metallicities than stellar mass or SFR selections.

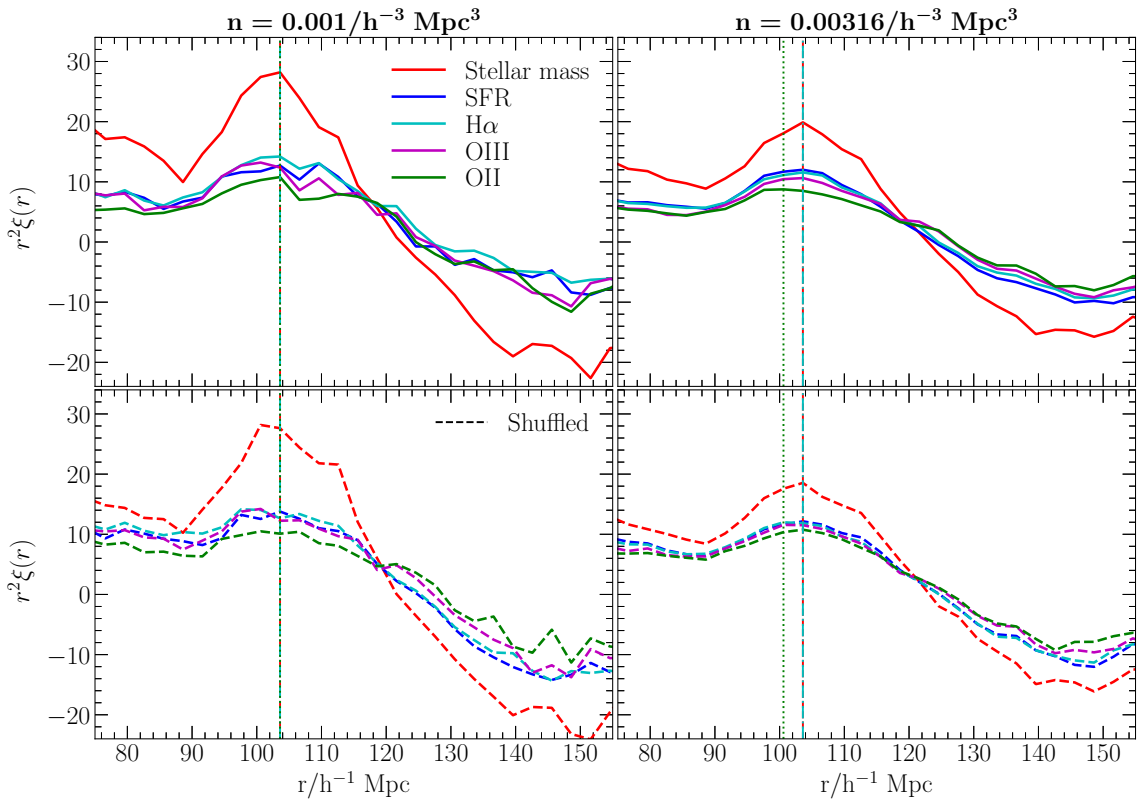

Figure 3.16: The Baryon Acoustic Oscillation (BAO) signals for the SAG samples (top) and their shuffled samples (bottom) of number density n = 0.001/h^{-3} Mpc^3 (left) and n = 0.00316/h^{-3} Mpc^3 (right). Colors correspond to the selection used. The lines are included to guide the eye around the peak of the BAO for the stellar mass, Hα and OII selections (all overlapped for the lowest density sample). Note that the BAO peak positions, for OII selected samples, is at lower scales than the other selections.

model it to prove the existence of assembly bias using OII selected galaxies from future surveys. However, it is still necessary to include the clustering uncertainties to claim that the scale shift is actually real.

We finally determine the β parameter, which depends on the logarithmic growth rate, and the bias parameter of each SAG sample. The former is a function of the matter density parameter, and the latter describes how biased are the galaxies to trace the underlying dark matter density field. The β parameter can be obtained from the ratio between the monopoles of the correlation functions in real and redshift space (Kaiser, 1986).

$$\xi_0(s) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)\xi(r) \tag{3.2}$$

Fig. 3.17 shows the β parameter as a function of scale for the SAG and ahuffled samples with two different number densities. This function is roughly constant for the stellar mass selection as expected from linear theory of perturbations, but it becomes highly non-constant for the OII and OIII cases. Indeed, we see a scale dependence for the shuffled samples, which indicates an effect from a non-constant large scale bias, as shown in Fig. 3.6.

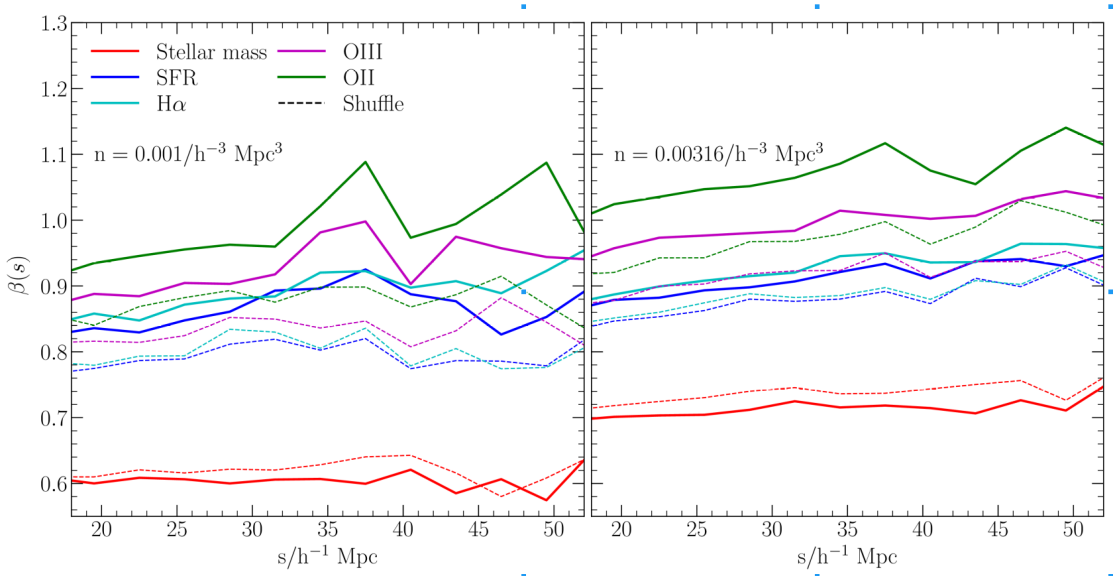

Figure 3.17: The evolution of the beta parameter as a function of scale-separation for n = 0.001/h^{-3} Mpc^3 (left) and n = 0.00316/h^{-3} Mpc^3 (right). Colors indicate the galaxy selection for the SAG samples (solid). The shuffled versions (dashed) for each sample are also shown. Note that the beta parameter is non-constant for ELG samples, which indicates a non-constant large scale bias.

Chapter 4

Conclusions

The next generation of surveys will measure the clustering of the galaxy distribution over a wide range of redshifts. Mock catalogues have proven to be important tools in preparation for this because of their multiple applications including error analysis, data interpretation and survey planning. SAMs are a physical approach to obtain such mocks, but sometimes their direct application to a simulation is not possible due to the limited resolution of the halo merger trees (Angulo et al., 2014) or the trees may not be available, as in the case of the Euclid flagship simulation (Potter et al., 2017). Even if the trees were available at the required resolution, the sheer number of halos in a giga-parsec side N-body simulation may preclude a direct calculation with a SAM.

The HOD model provides a simple yet efficient way to construct mock catalogues. The modelling consists of using a probability distribution to obtain the number of galaxies hosted in a halo of a particular mass. This simple method allows us to create large sets of mock catalogues for huge cosmological simulations. This is useful in the context of ELGs, as they are targets in current and coming surveys.

To determine the level of complexity needed to produce accurate mock catalogues, we test different HOD models. The 1-HOD uses the HOD of all galaxies making no distinction between centrals and satellites while the 2-HOD uses the HOD of these two components separately. The 4-HOD stores additional information about whether or not haloes host a central, and it constructs conditional HODs for satellites taking into account this information.

Because SAMs include assembly bias by construction, and in their simplest form HOD mocks

do not, we remove the assembly bias from the G13 SAM samples by shuffling the galaxy populations among haloes of the same mass creating the shuffled catalogue. This allows us to make a direct comparison between the clustering of our mocks and the SAMs from which we extract the HOD measurements. For example, we find that, for the intermediate galaxy density sample in the G13 SAM, the assembly bias affects the 2-halo term of the 2PCF of stellar mass selected galaxies increasing the amplitude by $\sim 12\%$. For the SFR selected galaxies, in contrast, the assembly bias suppresses the clustering by $\sim 4\%$. We also impose the standard NFW profile for satellites in the shuffled catalogue as is done for satellites in the HOD mock catalogues. We can then check the accuracy of the HOD models through a comparison between the 2PCFs of the HOD mocks and the shuffled-NFW catalogues.

The 2-HOD and 4-HOD produce the best mock catalogues as their 2PCFs are in close agreement with the clustering of the shuffled-NFW sample. We obtain the best results using a negative binomial distribution for the (conditional) HOD (see Eq. 1); in previous works, this was commonly considered to be a Poisson distribution. This is consistent with the subhalo HOD found in Boylan-Kolchin et al. (2010) using the Millennium-II simulation. Furthermore, we found that the assumption of this non-Poissonian HOD changes the galaxy clustering. Previously, Jiang & van den Bosch (2017) found a similar result using subhaloes from the Bolshoi, the MultiDark N-body simulations, and the SAM presented in Jiang & van den Bosch (2016).

The scatter of the HOD of satellites in G13 is reproduced by a negative binomial distribution up to halo masses of $M_h \lesssim 10^{13.5} M_\odot h^{-1}$. The galaxies in this halo mass range dominate the amplitude of the 2PCF. We quantify the departure from the Poisson distribution with the parameter β (see Eq. 2.3). We obtain the best clustering predictions for SFR selected samples using $\beta = 0.05$ and $\beta = 0.08$ for stellar mass selected samples. These correspond to negative binomial distributions slightly wider than Poisson. Because of the specific modelling of different SAMs, we expect that the best β values for each sample are model-dependent. For stellar mass selected samples, we find that the HOD scatter has a weak impact on clustering, making the addition of this additional parameter unnecessary in the context of mock catalogues.

The analysis of the HOD of satellites is important because the width of the distribution (determined by the β parameter) has a large impact on the one-halo term of the 2PCF of mock catalogues that emulate SFR-selected sample and ELG samples. If we consider the Poisson distribution for

the HOD of satellites ($\beta = 0$) the 2PCF of the mock catalogues is underestimated with respect to the clustering of the shuffled-NFW. In contrast, using the negative binomial distribution increases the amplitude of clustering in the one-halo regime. If we assume a value of β larger than the one present in the distribution of number of satellites, the clustering on small scales is further overestimated. We highlight the importance to perform a careful analysis of the satellite HOD if the HOD framework is used to produce mock catalogues, for ELGs or star-forming galaxies, following a particular model or observation.

The development of spectroscopic surveys, as the extended-Baryon Oscillation Spectroscopic Survey (eBOSS), allow to measure the galaxy density field and use it to constrain cosmology. In particular, emission line galaxies (ELGs) can be used to trace the underlying matter density field. Future experiments as the Dark Energy Spectroscopic Instrument (DESI) and the Euclid mission will detect ELGs making the previous analysis of these samples a necessary task to leverage the future galaxy catalogues.

We study the clustering of ELGs from different a semi-analytical model (SAG) implemented into the MultiDarkPlank 2 N-body simulation at $z \sim 1$ (similar to ELGs in eBOSS). The nebular emission is obtained from a photoionization code that uses the cold gas metallicity and instantaneous SFR of each galaxy as the input parameters. We define our galaxy samples according to their Hα, [OIII]5007 and [OII]$\lambda\lambda$3727 + 3729 emissions.

We measure the impact of assembly bias in the galaxy clustering, finding a clear scale-dependent signature for the [OIII] and [OII] selections. We analyze this in terms of the evolution of the large scale bias, that relates the clustering of galaxies with respect to the clustering of matter. We find that the bias is non-constant in large scales ($r \gtrsim 30$ Mpc h^{-1}) for the same selections mentioned above. This indicates that the large scale structure traced by these galaxies may contain some additional features that can potentially impact the measured clustering. Indeed, we find that that galaxy hosted by low mass haloes in underdense environments shows the scale-dependent assembly bias signature. Moreover, the cold gas metallicity of these galaxies tends to be lower than galaxies in other selections. Hence, a possible origin for this scale-dependence is that the [OII] and [OIII] selections also contain an environment selection.

We analyze the impact of this scale-dependence on the Baryon Acoustic Oscillations (BAO) scale and in the β parameter. For the [OII] selected samples, we find that the BAO peak position

is at smaller scales than in stellar mass or SFR selected samples. Moreover, the β parameters for the [OII] selections show a very steep relation with scale-separation, something not expected from a linear theory assumption. It is still necessary to include the error analysis for the clustering up to BAO scales and for the β measurement, to validate these new results on ELG clustering. If both results are real, the scale-dependence introduces systematic in the inferred cosmology constraints from the BAO peak and redshift space distortions.

Acknowledgments

This work was made possible by the efforts of Gerard Lemson and colleagues at the German Astronomical Virtual Observatory in setting up the Millennium Simulation database in Garching. The CosmoSim database used in this thesis is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark database was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064. The author gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) and the Partnership for Advanced Supercomputing in Europe (PRACE, www.prace-ri.eu) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de). The works presented in this thesis were supported by BASAL 170002, Fondecyt regular 1191813, and the NSF gran AST-1612085. These projects have received funding from the European UnionâĂŹs Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 734374. The calculations for these works were performed on the Geryon computer at the Center for Astro-Engineering UC, part of the BASAL PFB-06, which received additional funding from QUIMAL 130008 and Fondequip AIC-57 for upgrades.

Bibliography

Abbott T. M. C., et al., 2019, Physical Review Letters, 122, 171301

Alam S., et al., 2017, MNRAS, 470, 2617

Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, MNRAS, 399, 983

Angulo R. E., White S. D. M., Springel V., Henriques B., 2014, MNRAS, 442, 2131

Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, MNRAS, 480, 3978

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., 1999, MNRAS, 305, L21

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109

Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18

Benson A. J., 2010, Phys. Rep., 495, 33

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Blanton M., Cen R., Ostriker J. P., Strauss M. A., 1999, ApJ, 522, 590

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, MNRAS, 406, 896

Cochrane R. K., Best P. N., 2018, MNRAS, 480, 864

Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717

Contreras S., Zehavi I., Padilla N., Baugh C. M., Jiménez E., Lacerna I., 2019, MNRAS, 484, 1133

Cora S. A., 2006, MNRAS, 368, 1540

Cora S. A., et al., 2018, MNRAS, 479, 2

Croton D. J., et al., 2006, MNRAS, 365, 11

Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303

DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101

DeRose J., et al., 2019, ApJ, 875, 69

Delubac T., et al., 2017, MNRAS, 465, 1831

Dopita M. A., Sutherland R. S., 1995, ApJ, 455, 468

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Gargiulo I. D., et al., 2015, MNRAS, 446, 3820

Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, MNRAS, 426, 679

Gonzalez-Perez V., et al., 2018, MNRAS, 474, 4024

Grasshorn Gebhardt H. S., et al., 2019, ApJ, 876, 32

Groves B. A., Dopita M. A., Sutherland R. S., 2004, ApJS, 153, 9

Guo Q., et al., 2011, MNRAS, 413, 101

Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351

Guo Q., et al., 2016, MNRAS, 461, 3457

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373

Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848

Jiang F., van den Bosch F. C., 2017, MNRAS, 472, 657

Jiménez E., Contreras S., Padilla N., Zehavi I., Baugh C. M., Gonzalez-Perez V., 2019, arXiv e-prints, p. arXiv:1906.04298

Kaiser N., 1986, MNRAS, 222, 323

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

Kreckel K., Croxall K., Groves B., van de Weygaert R., Pogge R. W., 2015, ApJ, 798, L15

Lagos C. D. P., Cora S. A., Padilla N. D., 2008, MNRAS, 388, 587

Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193

Manera M., et al., 2013, MNRAS, 428, 1036

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Muñoz Arancibia A. M., Navarrete F. P., Padilla N. D., Cora S. A., Gawiser E., Kurczynski P., Ruiz A. N., 2015, MNRAS, 446, 2291

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19

Orsi Á. A., Angulo R. E., 2018, MNRAS, 475, 2530

Orsi Á., Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, MNRAS, 443, 799

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Planck Collaboration et al., 2014, A&A, 571, A16

Potter D., Stadel J., Teyssier R., 2017, Computational Astrophysics and Cosmology, 4, 2

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

Schaye J., et al., 2015, MNRAS, 446, 521

Scoccimarro R., Feldman H. A., Fry J. N., Frieman J. A., 2001, ApJ, 546, 652

Sinha M., Garrison L., 2017, Corrfunc: Blazing fast correlation functions on the CPU, Astrophysics Source Code Library (ascl:1703.003)

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Springel V., et al., 2005, Nature, 435, 629

Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. D. P., 2010, MNRAS, 408, 2008

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Wechsler R. H., Tinker J. L., 2018, preprint, (`arXiv:1804.03097`)

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

Wegner G., Grogin N. A., 2008, AJ, 136, 1

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Xu X., Zheng Z., 2018, arXiv e-prints, p. arXiv:1812.11210

Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

Zehavi I., et al., 2011, ApJ, 736, 59

Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, ApJ, 853, 84

Zehavi I., Kerby S. E., Contreras S., Jiménez E., Padilla N., Baugh C. M., 2019, arXiv e-prints, p. arXiv:1907.05424

Zheng Z., et al., 2005, ApJ, 633, 791