SEMI-ANALYTIC GALAXY FORMATION DURING THE EPOCH OF REIONISATION

by

Yisheng Qiu

ORCID: 0000-0002-7716-1094

A thesis submitted in total fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Science
School of Physics
THE UNIVERSITY OF MELBOURNE

October 2020

THE UNIVERSITY OF MELBOURNE

Abstract

Faculty of Science
School of Physics

Doctor of Philosophy

by Yisheng Qiu
ORCID: 0000-0002-7716-1094

Semi-analytic models play an important role in modelling the epoch of reionisation. This thesis presents three studies that are related to this topic. First, we measure clustering segregation with both UV-luminosity and stellar mass at $z \gtrsim 4$, which is then compared with predictions from the MERAXES semi-analytic model. Our results suggest that the dependence of clustering strength on UV-luminosity is stronger than stellar mass, indicating that compared with stellar mass, UV-luminosity is more tightly correlated with halo mass. Secondly, we investigate dust extinction in the early Universe. Our method utilises the MERAXES semi-analytic model to produce intrinsic galaxy luminosity and adopts parametric relations to estimate dust extinction. A novelty of our approach is that intrinsic luminosity and dust extinction are determined simultaneously by calibrating both galaxy formation and dust models against only UV observations. Our results suggest that there is a factor of two systematic error in the estimations of the cosmic star formation rate density based on the dust law in the local Universe. Finally, we present a method to augment N-body simulations using a Monte Carlo algorithm, which increases the mass resolution of the simulations. The results can be used by semi-analytic models of reionisation to overcome the challenge that convergent predictions of the reionisation history require both high mass resolution and large simulation volume. The effectiveness of our method is tested using a high resolution small volume N-body simulation.

Declaration of Authorship

I, Yisheng Qiu, declare that this thesis titled, 'Semi-analytic Galaxy Formation During the Epoch of Reionisation' and the work presented in it are my own. I confirm that:

- The thesis comprises only my original work towards the PhD except where indicated in the preface;
- due acknowledgement has been made in the text to all other material used; and
- the thesis is fewer than the maximum word limit in length, exclusive of tables, maps, bibliographies and appendices as approved by the Research Higher Degrees Committee.

Preface

This thesis comprises of my own work, except where explicitly mentioned in the text.

- The galaxy catalogue for the clustering measurements in Chapter 3 is provided by my collaborator, Pascal Oesch.
- The N-body simulation data used in Chapter 5 are provided by my collaborator, Pascal Elahi.

The publication status of the chapters is described as follows:

- Chapter 3 have been published as
  Qiu, Yisheng, Wyithe, J. Stuart B., Oesch, Pascal A., Mutch, Simon J., Qin, Yuxiang, Labbé, Ivo, Bouwens, Rychard J., Stefanon, Mauro, Illingworth, Garth D.
  Dependence of galaxy clustering on UV luminosity and stellar mass at $z \sim 4$-7
  Monthly Notices of the Royal Astronomical Society, Volume 481, Issue 4, p.4885-4894
- Chapter 4 have been published as
  Qiu, Yisheng, Mutch, Simon J., da Cunha, Elisabete, Poole, Gregory B., Wyithe, J. Stuart B.
  Dark-age reionization and galaxy formation simulation - XIX. Predictions of infrared excess and cosmic star formation rate density from UV observations
  Monthly Notices of the Royal Astronomical Society, Volume 489, Issue 1, p.1357-1372
- Chapter 5 was submitted to Monthly Notices of the Royal Astronomical Society as
  Yisheng Qiu, Simon J. Mutch, Pascal J. Elahi, J. Stuart B. Wyithe
  An efficient hybrid method to produce high resolution large volume dark matter simulations for semi-analytic models of reionisation

Acknowledgements

I wish to express my sincere gratitude to my supervisor, Stuart Wyithe, who provided invaluable support throughout my PhD. His guidance is essential in finalising my projects, which might be the most difficult part in the research. He showed impressive patience in reviewing and correcting my writing. He also encouraged and supported me to attend a variety of academic programs including training, conferences and an international visit.

I would like to pay my special regards to my co-supervisor, Simon Mutch. My works would have been impossible without the software developed by him. Thanks for his patience on both short and long discussions on my projects. He also helped me to solve many difficult technical issues in my research.

I am grateful to my international collaborator, Pascal Oesch. He provided accommodation and food during my one-month visit in the Geneve Observatory. The clustering project cannot be finished without his help. He also kindly wrote a reference letter for my job application.

I would like to thank my collaborators, Elisabete da Cunha and Pascal Elahi, who provided many ideas and suggestions on my projects.

I would like to thank my colleagues, Bradley Greig, Chuanwu Liu, Jeahong Park, Madeline Marshall and Yuxiang Qin for useful discussions.

Finally, thanks also go to the ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D). The centre organised several excellent academic programs, which broaden my knowledge in astronomy and make my student life more fruitful.

Contents

Abstract i

Declaration of Authorship ii

Preface iii

Acknowledgements iv

List of Figures viii

List of Tables x

Abbreviations xi

1 Introduction 1

1.1 Standard cosmology . . . 1
1.1.1 Friedmann-Robertson-Walker metric . . . 1
1.1.2 Redshift . . . 2
1.1.3 Friedmann equation . . . 2
1.2 Observations of high redshift galaxies . . . 3
1.2.1 Lyman-break galaxies . . . 3
1.2.2 One-point statistics . . . 4
1.2.3 Two-point statistics . . . 4
1.2.4 Colour magnitude relations . . . 5
1.3 Observations of cosmic reionisation . . . 5
1.3.1 Gunn-Peterson test . . . 5
1.3.2 Thomson scattering of CMB photons . . . 6
1.3.3 21cm power spectrum . . . 6
1.4 Analytic framework of dark matter halos . . . 7
1.4.1 Formation of dark matter halos . . . 8
1.4.2 Halo mass functions . . . 9
1.4.3 Conditional mass functions . . . 10
1.4.4 Halo clustering . . . 10
1.5 Galaxy formation simulations . . . 11
1.5.1 N-body simulations . . . 11

1.5.2 Semi-analytic models . . . 12
1.5.3 Hydrodynamical simulations . . . 13
1.6 Cosmic reionisation simulations . . . 13
1.7 Thesis structure . . . 14

2 The Meraxes semi-analytic models 15
2.1 Gas infall . . . 15
2.2 Cooling . . . 16
2.3 Star formation . . . 16
2.4 Supernova feedback . . . 17
2.5 Mass recycling and metal enrichment . . . 19
2.6 Reionisation feedback . . . 19

3 Clustering segregation with stellar mass and UV-luminosity 21
3.1 Introduction . . . 21
3.2 Measuring clustering segregation . . . 22
3.2.1 Observational data . . . 22
3.2.2 Estimating the angular correlation function . . . 26
3.2.3 Estimating the correlation length and bias . . . 28
3.2.4 Results and discussion . . . 30
3.3 Comparison with Meraxes . . . 32
3.3.1 Model overview . . . 32
3.3.2 Results . . . 35
3.4 Summary . . . 36

4 Dust extinction at high redshifts 38
4.1 Introduction . . . 38
4.2 Updates to Meraxes . . . 40
4.3 Dust models . . . 41
4.3.1 Star formation rate model . . . 41
4.3.2 Dust-to-gas ratio model . . . 42
4.3.3 Gas column density model . . . 42
4.4 Synthetic spectral energy distributions . . . 43
4.5 Calibration . . . 46
4.6 Fitting results . . . 52
4.7 Infrared excess to UV continuum slope relation . . . 57
4.7.1 Reddening slope . . . 60
4.7.2 Intrinsic scatter . . . 60
4.8 Cosmic star formation density . . . 61
4.9 Summary . . . 63

5 Extending mass resolution of N-body simulations 65
5.1 Introduction . . . 65
5.2 The Genesis N-body simulations . . . 66
5.3 Augmenting N-body merger trees . . . 67
5.3.1 Generating Monte Carlo trees . . . 67
5.3.2 Augmentation algorithm . . . 69
5.3.3 Fixing original subhalo trees . . . 73

5.3.4 Identifying the complete halo population . . . 73
5.3.5 Applying to N-body simulations . . . 74
5.4 Assigning halo positions . . . 76
5.4.1 Populating halo positions . . . 76
5.4.2 Evolving halo positions . . . 77
5.5 Spin parameters . . . 81
5.6 Application to Meraxes . . . 82
5.6.1 Galaxy properties . . . 84
5.6.2 Reionisation histories . . . 85
5.7 Summary . . . 86

6 Summary **88**

List of Figures

1.1 Illustration of LBG selection at $z \sim 4$. . . . 3
1.2 Time evolution of 21cm brightness temperature. . . . 8
1.3 Visualisation of the dark matter density field predicted by the Millennium simulation. . . . 12

3.1 Scatter plots of the $M_\star - M_{\rm UV}$ relations. . . . 24
3.2 Measured ACFs and their best-fit power laws $A_\omega(\theta/1'')^{-0.6}$. . . . 25
3.3 Measured galaxy biases as a function of mean stellar mass and mean UV flux. . . . 28
3.4 Results of a straight line fit of measured biases over all redshifts. . . . 29
3.5 Example redshift distribution of observed and selected model galaxies in the $z \sim 4$ LBG sample. . . . 33
3.6 Comparison between observed and model predicted ACFs at $z \sim 4$. . . . 34

4.1 Best-fit luminosity functions (LFs) and colour-magnitude relations (CMRs). 44
4.2 Comparison of the marginalised distributions of galaxy formation parameters among the three different dust models. . . . 46
4.3 Posterior distribution of the galaxy and dust parameters for MERAXES with a star formation rate dependent (SFR) dust model. . . . 49
4.4 Posterior distribution of the galaxy and dust parameters for MERAXES with a dust-to-gas ratio (DTG) dependent dust model. . . . 50
4.5 Posterior distribution of the galaxy and dust parameters for MERAXES with a gas column density (GCD) dependent dust model. . . . 51
4.6 Correlations among the supernova energy coupling efficiency ϵ_0, galaxy property scaling of the dust relation $\gamma_{\rm SFR,\,DTG,\,GCD}$ and the reddening slope n. . . . 52
4.7 Effects of varying the mass loading factor η_0 and the supernova energy coupling efficiency ϵ_0 on the intrinsic UV luminosity function. . . . 53
4.8 Effects of varying the galaxy property scaling of the dust relation $\gamma_{\rm SFR,\,DTG,\,GCD}$ on the dust-attenuated UV luminosity functions and colour magnitude relations. . . . 54
4.9 Redshift evolution of the mass metallicity relation. . . . 57
4.10 Predicted infrared excess (IRX) - UV continuum slope β relations. . . . 58
4.11 Predicted infrared excess (IRX) - UV continuum slope β relations as functions of stellar mass (left panels) and specific star formation rate (sSFR) (right panels) at $z \sim 5$. . . . 59
4.12 Predicted cosmic star formation rate density (SFRD) at $z \sim 4 - 7$. . . . 62

5.1 Fitting results of the calibration for the Parkinson et al. (2008) algorithm. 68

5.2 Schematic diagram of augmenting N-body halo merge trees. 73
5.3 Comparisons of the conditional functions, defined by $df_{\rm CMF}/d\ln M_1$, of N-body and augmented merger trees. 74
5.4 Halo mass functions of extended halo catalogues. 75
5.5 Comparison of two-point correlation functions produced using the random sampling method and estimated from N-body simulations. 78
5.6 Comparison of two-point correlation functions produced using the evolving method and estimated from N-body simulations. 78
5.7 Peculiar velocity distributions of N-body and Monte Carlo halos. 79
5.8 Spin distributions of N-body and Monte Carlo halos. 81
5.9 Stellar mass functions and satellite fractions predicted by the MERAXES semi-analytic model. 82
5.10 Star formation rate functions and satellite fractions as a function of star formation rate predicted by the MERAXES semi-analytic model. 82
5.11 Star formation rate density and volume-weighted neutral fractions predicted by the MERAXES semi-analytic model. 83
5.12 Effect of cosmic variance on the reionisation history. 83

List of Tables

3.1 Best-fit parameters of the $M_\star - M_{\rm UV}$ relations 26
3.2 Summary of clustering measurements at $z \sim 4 - 7$. 30

4.1 The five windows selected from Calzetti et al. (1994) to fit UV slopes for the on-the-fly calibrations. 44
4.2 Summary of free galaxy and dust parameters. 45
4.3 Tabular data of predicted cosmic star formation rate density (SFRD) for the three different dust models. 63

5.1 Parameters of the Monte Carlo tree algorithm. 67
5.2 Parameters of the tree augmentation algorithm. 72
5.3 Information on halo catalogues used in this work. 72

Abbreviations

ACF	**A**ngular **C**orrelation **F**unction
CMB	**C**osmic **M**icrowave **B**ackground
CMR	**C**olour **M**agnitude **R**elation
LBG	**L**yman **B**reak **G**alaxy
LF	**L**uminosity **F**unction
EoR	**E**poch **o**f **R**ionisation
EPS	**E**xtended **P**ress–**S**chechter
HERA	**H**ydrogen **E**poch **R**eionization **A**rray
HST	**H**ubble **S**pace **T**elescope
IGM	**I**nter-**g**alatic **M**edia
IMF	**I**nitial **M**ass **F**unction
IRX	**I**fra**r**ed e**x**cess
ISM	**I**nter-**s**tellar **M**edia
JWST	**J**ames **W**ebb **S**pace **T**elescope
MWA	**M**urchison **W**idefield **A**rray
SED	**S**pectral **E**nergy **D**istribution
SKA	**S**quare **K**ilometre **A**rray
SMF	**S**tellar **M**ass **F**unction
SSP	**S**imple **S**tellar **P**opulation
UVB	**U**ltra**v**iolet **B**ackground

Chapter 1

Introduction

In the past decades, the measurements of cosmology parameters are becoming more and more precise owning to routine observations of the cosmic microwave background (CMB) and type Ia supernovae. The observations of the CMB also provide the initial conditions for subsequent structure formation. Accordingly, many interests are extended to the next phase transition of the Universe. The first generation of stars and galaxies formed in denser regions in the early Universe, and started to ionise the neutral hydrogen in surrounding regions. This is known as the epoch of reionisation (EoR). This thesis focuses on several topics related to the modelling of the EoR. To begin with, this chapter provides an introduction of background knowledge.

1.1 Standard cosmology

1.1.1 Friedmann-Robertson-Walker metric

The evolution of the Universe can be described by a space-time metric. The form of the metric can be dramatically simplified by the fundamental principle of cosmology, which states that the Universe is homogeneous and isotropic on sufficient large scales. This symmetry leads to the Friedmann-Robertson-Walker metric:

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \tag{1.1}$$

where c is the speed of light and k can be -1, 0, 1, corresponding to an open, flat and close universe respectively. A mathematically derviation of the metric can be found in Weinberg (1973). Current observations suggest that $k = 0$ (Planck Collaboration et al., 2016). The function $a(t)$ is known as the scale factor, which describes the expansion of the Universe. The equation that describes $a(t)$ can be obtained by substituting the Friedmann-Robertson-Walker metric into the Einstein equation of general relativity.

1.1.2 Redshift

Given the metric in Equation 1.1, consider a photon that is emitted at t with wavelength λ. It can be shown that when the photon is observed at $t_{\rm obs}$, its wavelength follows (see e.g. Landau & Lifshitz, 1975; Mo et al., 2010)

$$\lambda_{\rm obs} = \frac{1+z}{1+z_{\rm obs}}\lambda, \tag{1.2}$$

with

$$1+z = \frac{1}{a(t)}, \quad 1+z_{\rm obs} = \frac{1}{a(t_{\rm obs})}. \tag{1.3}$$

We choose $a(t_{\rm obs}) = 1$ by convention. Accordingly, in an expanding Universe, when the photon is observed, its wavelength is increased by $1+z$. This is known as cosmological redshift.

1.1.3 Friedmann equation

The Friedmann equation describes the time evolution of $a(t)$, or equivalently $z(t)$, and can be obtained by substituting the Friedmann-Robertson-Walker metric into the Einstein equation of general relativity. If we are only interested in studying matter in the Universe and assume that the Universe is flat, the Friedmann equation takes the form

$$H(z) = H_0\sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}, \tag{1.4}$$

with

$$H(z) = -\frac{1}{1+z}\frac{dz}{dt}, \tag{1.5}$$

where H_0 is the Hubble constant, $\Omega_{\rm m}$ is the density fraction of matter, and Ω_Λ is the density fraction of dark energy. Planck Collaboration et al. (2016) measured from the

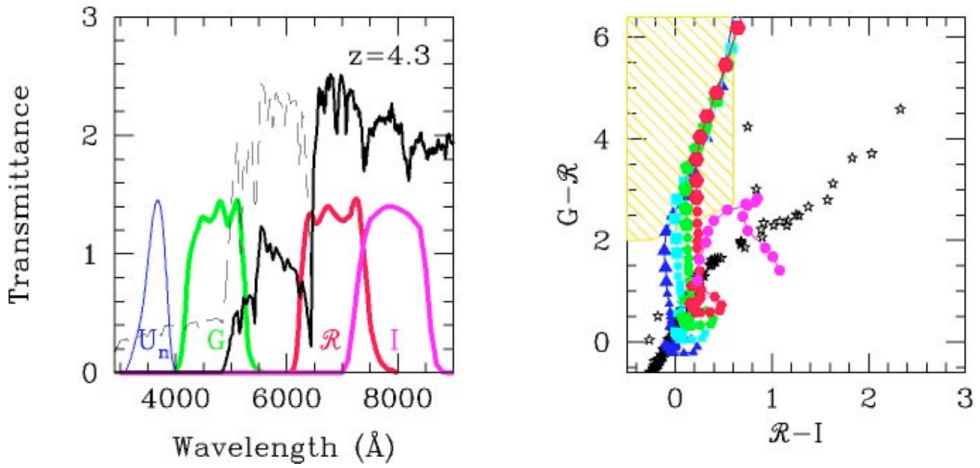


FIGURE 1.1: Illustration of LBG selection at $z \sim 4$. Left panel shows the Lyman break from a galaxy spectrum, which is identified using the G and R filters. In right panel, galaxies within the yellow shaded region are selected as LBGs. The figure is taken from Giavalisco (2002).

cosmic wave background (CMB) that $H_0 = 67.8\,\mathrm{km/s/Mpc}$, $\Omega_\mathrm{m} = 0.308$ and $\Omega_\Lambda = 0.692$. Matter in the Universe is classified into baryonic and dark matter. Planck Collaboration et al. (2016) also measured a baryonic fraction $\Omega_\mathrm{b} = 0.0484$. Therefore, the Universe is mostly comprised of dark energy and dark matter. For the reason, the standard cosmology is known as the lambda cold dark matter (ΛCDM) model.

1.2 Observations of high redshift galaxies

1.2.1 Lyman-break galaxies

Most high redshift galaxies are identified using the Lyman-break techniques (Steidel et al., 1996). At $z \gtrsim 3$, the IGM is opaque to photons with wavelength shorter than 1216Å due to the Lyman absorption, resulting in a sharp drop in galaxy spectra. This feature can be observed in the optical range due to cosmological redshift and can be used to select galaxies. Figure 1.1 gives an example of the LBG selection at $z \sim 4$. The left panel illustrates the Lyman break, which is identified using the G and R filters. In the right panel, galaxies within the yellow shaded region are selected as LBGs. A review of LBGs can be found in Giavalisco (2002).

1.2.2 One-point statistics

One-point statistics are number count of galaxies as a function of a certain property. At $z \gtrsim 3$, rest-frame UV fluxes can be effectively observed by optical telescopes due to cosmological redshift. Hence, UV observations are fundamental in the early Universe. A robust observable of LGBs is the rest-frame UV luminosity function (LF) (van der Burg et al., 2010; Bouwens et al., 2015; Finkelstein et al., 2015; Ono et al., 2018), which is defined by the number density of galaxies as a function of UV magnitude. Moreover, stellar mass, which might be the most important property of a galaxy, can be estimated by SED fittings (see Conroy, 2013, for a review). Several studies have also measured the stellar mass function (SMF) for LBGs (Duncan et al., 2014; Song et al., 2016). Infrared data are critical in the measurements. Since observations of infrared data are very challenging at high redshifts, the uncertainties of SMFs are much larger than UV LFs. One-point statistics are fundamental observations of galaxies, which are compared with simulations in the first place.

1.2.3 Two-point statistics

Two-point statistics describe the spatial distribution of galaxies. The two-point correlation function can be estimated by

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1, \tag{1.6}$$

where $DD(r)$ is the probability of finding galaxies pairs at separation r, and $RR(r)$ is the probability of finding pairs of uniformly distributed samples at separation r. The dependence of the two-point correlation function on certain galaxy properties is known as clustering segregation. Such trend with UV luminosity has been detected for LBGs up to $z \sim 7$ (Barone-Nugent et al., 2014; Harikane et al., 2016). Clustering segregation of LBGs has also been observed with stellar mass (Ishikawa et al., 2017; Driver et al., 2018). Observations of clustering segregation reveal an underlying correlation between galaxy properties and halo mass. This point will be discussed in Section 1.4.4.

1.2.4 Colour magnitude relations

An additional observable of LBGs is the UV continuum slope. At wavelengths roughly between 1300Å and 2600Å, the galaxy spectrum can be described by a power law, and the power law slope is referred to as the UV continuum slope. This quantity is found to be correlated with UV magnitude. This correlation is known as the colour magnitude relation (CMR), which have been observed at $z \gtrsim 3$ (Finkelstein et al., 2012; Bouwens et al., 2014). UV luminosity is an indicator of star formation, which, however, is heavily extincted by interstellar dust. The CMR can be combined with the Meurer et al. (1999) infrared excess (IRX) to UV continuum slope relation to estimate the dust-corrected star formation rate (e.g. Bouwens et al., 2015; Mason et al., 2015; Liu et al., 2016).

1.3 Observations of cosmic reionisation

Shortly after the big bang (at $z \sim 1000$), the Universe cools down due to the cosmological expansion so that hydrogen atoms can form. This phase transition is known as recombination. Photons emitted by the combination of a proton and an electron during this era can be observed through the cosmic microwave background (CMB). After recombination, ordinary matter in the Universe is dominated by neutral hydrogen, mixed with a small fraction of helium.

The subsequent phase transition of the Universe, i.e. the epoch of reionisation (EoR), started when neutral hydrogen is ionised by the surrounding light sources. Observations of the CMB have found fluctuations in the density of the Universe. Current theory suggests that matter in denser regions can collapse due to the gravitational force, which creates the environment to form stars and galaxies. These light sources ionise the surrounding media. The following subsections introduce three probes to observe this phase transition.

1.3.1 Gunn-Peterson test

The Gunn-Peterson test measures the optical depth of Lyman alpha absorption using quasars. This approach was initially proposed by Gunn & Peterson (1965). As mentioned in Section 1.2.1, Lyman alpha photons can be effectively absorbed by the neutral

hydrogen in the IGM. The optical depth of Lyman alpha photons can be related to the density of neutral hydrogen in the IGM by

$$\tau_{\rm GP} = \frac{\pi e^2}{m_e c} f_\alpha \lambda_\alpha \frac{n_{\rm HI}}{H(z)}, \tag{1.7}$$

where c is the speed of light, e is the electron charge, m_e is the electron mass, and $H(z)$ is given by Equation 1.4. The oscillator strength and wavelength of Lyman alpha transition are denoted as f_α and λ_α respectively. Quasars are much more luminous than galaxies and have a well-behaved continuum spectrum, which provides a nice probe for the optical depth of Lyman alpha photons. Current observations on high redshift quasars suggest a very small $\tau_{\rm GP}$ at $z \lesssim 6$ (Fan et al., 2006), implying the end of reionisation at this time.

1.3.2 Thomson scattering of CMB photons

CMB photons can interact with free electrons emitted during the epoch of reionisation. This interaction reduces fluctuations of CMB and polarises the CMB photons. The Thomson scattering optical depth along the line-of-sight can be written as

$$\tau_e = \sigma_{\rm T} \int_0^z \frac{n_e(z) c dz}{(1+z) H(z)}, \tag{1.8}$$

where c is the speed of the light, $\sigma_{\rm T}$ is the Thomson scattering cross section, and $H(z)$ is given by Equation 1.4. The electron density as a function of redshift $n_e(z)$ is related the neutral fraction using a particular reionisation model, and measurements of $\tau_{\rm e}$ provide constraints on the assumed model. For instance, Robertson et al. (2015) suggested that the epoch of reionisation occurs at $6 \lesssim z \lesssim 10$ assuming an analytic reionisation model and using the joint constraints of the Thomson scattering optical depth measured by Planck Collaboration et al. (2016) and the cosmic star formation rate density estimated by Madau & Dickinson (2014).

1.3.3 21cm power spectrum

The 21cm power spectrum provides a more informative probe of cosmic reionisation, which can measure ionising structures. The 21cm line is emitted due to the spin flip

of the hyperfine structure of hydrogen atoms. This wavelength is in the Rayleigh-Jeans limit compared with the black body radiation temperature of the CMB. Therefore, we can represent the 21cm intensity I_ν using the brightness temperature T_b by

$$I_\nu = \frac{2k_\mathrm{B}\nu^2 T_\mathrm{B}(\nu)}{c^2}, \tag{1.9}$$

where c is the speed of light and k_B is the Boltzmann constant. 21cm experiments aim to measure the fluctuation of the brightness temperature, which can be expressed by

$$\delta T_\mathrm{B} = \frac{T_\mathrm{S} - T_\gamma}{1+z}(1 - e^{-\tau_\nu}) \tag{1.10}$$

$$\approx 27\mathrm{mK} \times x_\mathrm{HI}(1+\delta)\left(1 - \frac{T_\gamma}{T_\mathrm{S}}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_m h^2}\right)^{1/2}\left(\frac{\Omega_b h^2}{0.023}\right), \tag{1.11}$$

where T_S is the spin temperature, T_γ is the CMB temperature, and τ_ν is the 21cm optical depth. The above equation suggests that δT_B depends on three important quantities, i.e. the neutral fraction x_HI, the density contrast of the underlying matter δ (see Equation 1.12), and cosmology parameters including h, Ω_m and Ω_b. Therefore, the measurement of δT_B put constraints on these quantities and the underlying physics. Figure 1.2 shows the evolution of the 21cm brightness temperature with respect to cosmic time. However, observing the brightness temperature fluctuation is extremely challenging, since the signal is very weak and is heavily contaminated by extragalactic radio sources and galactic foreground. Current observations from the Murchison Widefield Array (MWA) can put upper limits on the 21cm power spectrum (Barry et al., 2019). Robust measurements of the 21cm signal are expected to be obtained from future instruments, e.g. the Square Kilometre Array (SKA) and the Hydrogen Epoch of Reionization Array (HERA).

1.4 Analytic framework of dark matter halos

As introduced in Section 1.1.3, most matter in the Universe is dark matter. On small scales, dark matter is not uniformly distributed. Due to gravitational instability, dark matter can collapse and form a compact object named as dark matter halos. These halos are assumed to host galaxies. Small halos can merger together and form larger halos. Thus, the merger history of dark matter halos forms a tree structure, which is known as dark halo merger trees. The formation of such tree structures provides the

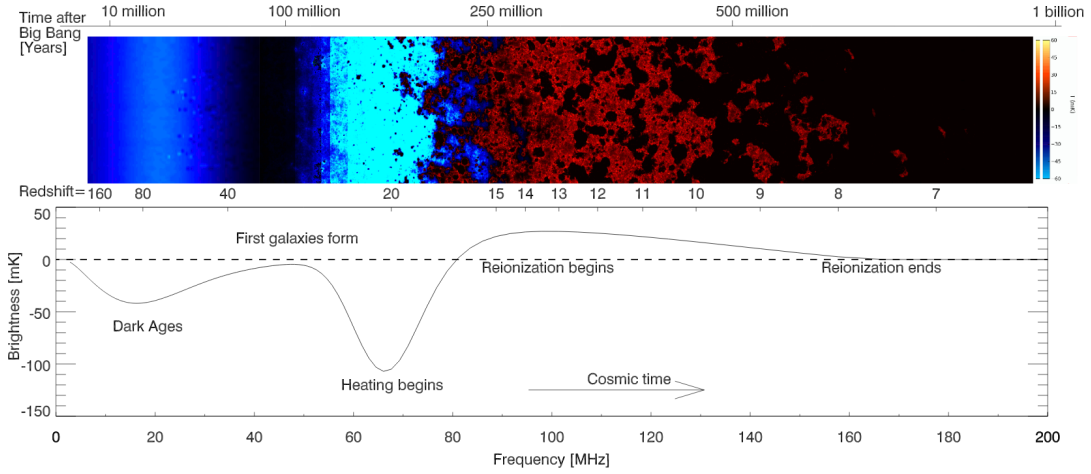


FIGURE 1.2: Time evolution of 21cm brightness temperature. This figure is taken from Pritchard & Loeb (2012).

basic picture of how matter evolves in the Universe. This section briefly reviews the analytic predictions for dark matter halos. These halo properties can better be studied by numerical simulations, which will be introduced in Section 1.5.1.

1.4.1 Formation of dark matter halos

To study formation of dark matter halos, we start by expressing the dark matter density field as

$$\delta_{\rm DM}(\vec{x},t) = \frac{\rho_{\rm DM}(\vec{x},t)}{\bar{\rho}_{\rm DM}} - 1, \tag{1.12}$$

where $\rho_{\rm DM}(\vec{x},t)$ and $\bar{\rho}_{\rm DM}$ are the density and mean density of dark matter. We can treat dark matter as collisionless ideal gas and describe its evolution using Euler's equations of fluid mechanics. In the linear regime, the equations have the following growth mode solution:

$$\delta_{\rm DM}(\vec{x},t) = \frac{D(t')}{D(t)}\delta_{\rm DM}(\vec{x},t'), \tag{1.13}$$

with

$$D(t) = \frac{H(z)\int_z^\infty dz'(1+z')/H^3(z')}{H_0\int_0^\infty dz'(1+z')/H^3(z')}, \tag{1.14}$$

where $H(z)$ is given by Equation 1.4.

The dark matter density field of a finite region can be obtained by convolving the the field using a window function $W(\vec{x}; R)$:

$$\delta(\vec{x}, t; R) = \int W(|\vec{x} - \vec{r}|; R)\delta_{\rm DM}(\vec{x}, t)d^3\vec{r}, \tag{1.15}$$

where R is the characterised size of the smoothing region. Since convolution is a linear operation, the linear growth factor given by Equation 1.14 can also be applied to the smoothed density field. For a spherical collapse model, it can be shown that a region can collapse and form a dark matter halo once the corresponding (linear) smoothed density field exceeds a critical value:

$$\delta_c(t) = 1.686/D(t) \tag{1.16}$$

1.4.2 Halo mass functions

By assuming that the dark matter density field follows a Gaussian distribution, Press & Schechter (1974) found that the abundance dark matter halos can be described by

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}}\frac{\rho_0}{M}\frac{\delta_c}{\sigma}\exp\left(-\frac{\delta_c^2}{2\sigma^2}\right)\left|\frac{d\ln\sigma}{dM}\right|, \tag{1.17}$$

where ρ_0 is the mean mass density of the Universe and δ_c is given by Equation 1.16. The mass variance of collapsed regions σ can be calculated by

$$M = \frac{4}{3}\pi\rho_0 R^3, \tag{1.18}$$

$$\sigma(R) = \frac{1}{2\pi^2}\int_0^\infty dk\, k^2 P(k) W^2(kR), \tag{1.19}$$

$$W(kR) = \frac{3[\sin(kR) - kR\cos(kR)]}{(kR)^3}, \tag{1.20}$$

where $P(k)$ is the power spectrum of the CMB. In the derivation of Equation 1.17, Press & Schechter (1974) manually increased dn/dM by a factor of two to resolve the cloud-in-cloud problem. Bond et al. (1991) provided a more elegant derivation, which is known as the extended Press-Schechter (EPS) formalism (see also Bower, 1991; Lacey & Cole, 1993). Comparison with numerical simulation suggests that the Press-Schechter halo mass function (given by Equation 1.17) is only slightly underestimated and overestimated at high and low mass ends respectively. A more consistent halo mass function can be obtained by assuming a ellipsoidal collapse model (Sheth et al., 2001).

1.4.3 Conditional mass functions

The EPS formalism can predict the the conditional mass function of dark matter halo, defined by the mass fraction (M_1/M_2) distribution of halos at z_1 with mass M_1 that will merger into a halo at z_2 with mass M_2. It is expressed by

$$\frac{df_{\rm CMF}}{dM_1} = \sqrt{\frac{2}{\pi}} \frac{\sigma_1^2(\delta_1 - \delta_2)}{(\sigma_1^2 - \sigma_2^2)^{3/2}} \exp\left[-\frac{1}{2}\frac{(\delta_1 - \delta_2)}{(\sigma_1^2 - \sigma_2^2)}\right] \left|\frac{d\ln\sigma_1}{dM_1}\right|, \tag{1.21}$$

where $\sigma_1 = \sigma(M_1)$, $\sigma_2 = \sigma(M_2)$, $\delta_1 = \delta_c(z_1)$, and $\delta_2 = \delta_c(z_2)$. The conditional mass function can also be applied to a region instead of a halo at z_2, in which case δ_2 should be replaced by the smoothed density field $\delta(\vec{x}, t; R)$. For a sufficient large region where $\delta(\vec{x}, t; R) \to 0$ and $\sigma_2 \to 0$, the conditional mass function can be reduced to the halo mass function given by Equation 1.17. Moreover, based on the conditional mass function, many Monte Carlo algorithms have been proposed to generate halo merger trees (e.g. Neistein & Dekel, 2008; Parkinson et al., 2008; Zhang et al., 2008), which can be combined with semi-analytic models (see Section 1.5.2) to study galaxy formation.

1.4.4 Halo clustering

An application of the conditional mass function (given by Equation 1.21) is to predict large scale clustering of halos. The halo density field of a spherical region with radius R_0 can be written as

$$\delta_{\rm h} = \frac{M_0}{M_1}\frac{df_{\rm CMF}}{dM_1}\left(\frac{dn}{dM}V_0\right)^{-1} - 1, \tag{1.22}$$

where M_0 and V_0 are the mass and volume of the spherical region. When $M_0 \gg M_1$, we can apply the Taylor expansion to the conditional mass function and obtain the following simple relation (Mo & White, 1996)

$$\delta_{\rm h} = b(M_1)\delta_{\rm DM}, \tag{1.23}$$

with

$$b(M_1) = \frac{\delta_1^2/\sigma_1^2 - 1}{\delta_1}, \tag{1.24}$$

where b is known as the halo bias. It follows that the two-point correlation function of dark matter halos $\xi_{\rm h}(r)$ (defined by Equation 1.6 with galaxy pairs replaced by halo

pairs) can be related to the correlation function of dark matter $\xi_{\rm DM}(r)$ by

$$\xi_{\rm h}(r) = b^2(M)\xi_{\rm DM}(r). \tag{1.25}$$

While $\xi_{\rm DM}(r)$ can be measured from the CMB, $\xi_{\rm h}(r)$ can be estimated from galaxies using a halo occupation model (e.g. Lee et al., 2006). Accordingly, Equation 1.25 provides a way to infer host halo mass of galaxies.

1.5 Galaxy formation simulations

There are three popular methods to study galaxy formation numerically, i.e. N-body simulations, semi-analytic models and hydrodynamical simulations. They approach the problem from different angles. This section provides a brief introduction to each of them.

1.5.1 N-body simulations

N-body simulations aim to evolve the gravitational field of dark matter. Although the nature of dark matter is still not well understood, modelling the dynamical evolution of dark matter is a relatively simple problem in astronomy, since only the gravitational force is involved. We are interested in formation of dark matter halos, which correspond to sharp peaks in the density field. Such structures can be well captured by representing the gravitational field using N particles. The problem then reduces to solve the equations of motion of these N particles. For the brute force method, The computational complexity to calculate the gravitational force of N particles scales as $O(N^2)$, which is the most time-consuming part of N-body simulations. Many algorithms have been proposed to reduce the computational cost, e.g. the tree algorithm (Barnes & Hut, 1986), the particle-mesh algorithm (Bouchet & Kandrup, 1985), and the particle-particle-particle-mesh algorithm (Efstathiou et al., 1985). With the help of modern supercomputers, a simulation with $\gtrsim 10^{10}$ particles is routinely accessible (e.g Springel et al., 2005; Watson et al., 2013; Poole et al., 2016). Figure 1.3 visualises the dark matter density field predicted by the Millennium simulation (Springel et al., 2005), with bright colours representing high density regions.

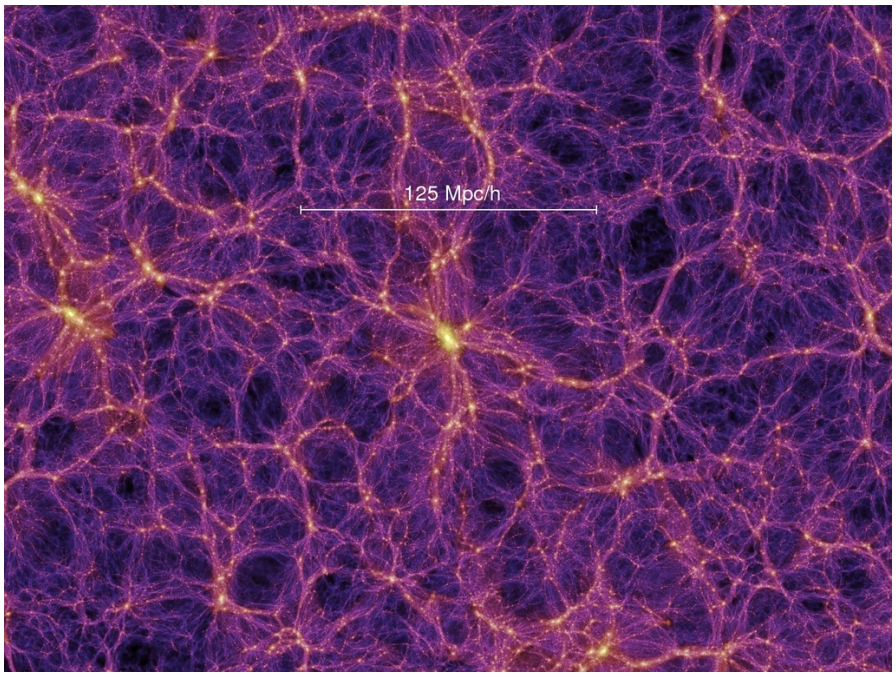


FIGURE 1.3: Visualisation of the dark matter density field predicted by the Millennium simulation (Springel et al., 2005). Bright colours correspond to high density regions.

1.5.2 Semi-analytic models

Semi-analytic models implement baryonic processes such as star formation based on pre-calculated merger histories of dark matter halos. Such treatment is motivated by the observational fact that the Universe is dominated by dark matter and dark energy. Halo merger trees either constructed from N-body simulations or generated using Monte Carlo algorithms can be used as input of semi-analytic models. The accretion of baryonic matter into dark matter halos can be estimated using a baryonic fraction. While the simplest approach is to adopt the global baryonic fraction measured from the CMB, studies suggested that the baryonic fraction may depend on multiple effects (e.g Sawala et al., 2013; Sobacchi & Mesinger, 2013; Qin et al., 2019b). In semi-analytic models, a galaxy is comprised of hot gas, cold gas and stars. Detailed models may also include disk, bulge and black hole components. A set of ordinary differential equations are proposed to model the baryonic processes that lead to mass transfer among these components, e.g. gas cooling, star formation, supernovae feedback. Since these processes are still not well understood, these equations typically contain many free parameters that need to be calibrated against observations. Studies have found that semi-analytic model can fit many observations, e.g. the SMF (Henriques et al., 2013, 2015; Cora et al., 2018) and the UV LF (Somerville et al., 2012; Cousin et al., 2019).

1.5.3 Hydrodynamical simulations

Hydrodynamical simulations evolve dark matter and baryonic matter simultaneously. In these simulations, astrophysical gases are treated as ideal gases that are governed by Euler's equations of fluid mechanics. Both particle-based and mesh-based numerical techniques are proposed to solve the hydrodynamical equations (e.g Gingold & Monaghan, 1977; Berger & Colella, 1989). Compared with semi-analytic models, hydrodynamical simulations require fewer assumptions on galaxy and dark matter halo structures, and provide much more self-consistent treatment for processes that need to be spatially resolved, e.g. galaxies mergers and supernova driven winds. However, these advantages come with a significant increase in computational cost. Large hydrodynamical simulations include ILLUSTRIS (van der Burg et al., 2010), EAGLE (Schaye et al., 2015), and BLUETIDES (Feng et al., 2016), which succeeded in reproducing a wide range of observations.

1.6 Cosmic reionisation simulations

Cosmic reionisation simulations aim to predict the ionising structure of the IGM, which can then be compared with the observations introduced in Section 1.3. These simulations include two key components, i.e. producing ionising sources and modelling radiative transfer in the IGM. The former can be realised using analytic models (Wyithe & Loeb, 2003, 2013), semi-analytic models (Ma et al., 2016; Yung et al., 2020) and hydrodynamical simulations (Wise et al., 2014; Ocvirk et al., 2016; Rosdahl et al., 2018). Alternatively, one can also assign ionising sources to dark matter halos in N-body simulations using a mass-to-light ratio (e.g. Iliev et al., 2006, 2014). The ionising structure of the IGM can be estimated using radiative transfer simulations, which have three classes: moments, Monte Carlo, and ray-tracing methods (see Trac & Gnedin, 2011, for a review). These simulations are computationally expensive. An alternative approach was recognised by Mesinger & Furlanetto (2007), which uses the excursion set theory to model ionising bubbles.

1.7 Thesis structure

The primary focus of this thesis is to explore applications of semi-analytic galaxy formation models during the EoR. This thesis is organised as follows:

- Chapter 2 reviews the MERAXES semi-analytic model, which is used throughout the paper.
- Chapter 3 measures clustering segregation with UV-luminosity and stellar mass at $z \gtrsim 4$. The measurements are then compared with predictions from MERAXES to study the underlying correlation between halo mass and these galaxy properties.
- Chapter 4 introduces an approach to estimate dust extinction in the early Universe by exploring the parameter space of MERAXES using constraints of only UV observations. The approach also provides self-consistent predictions of IRX.
- Chapter 5 presents a method to extend the mass resolution of N-body simulations using a Monte Carlo algorithm. The resulting halo catalogues can be used by semi-analytic models in order to resolve a complete population of faint galaxies in large volumes, which is critical to obtain convergent predictions for cosmic reionisation.
- Chapter 6 provides a summary of this thesis.

Chapter 2

The Meraxes semi-analytic models

In the studies of the following chapters, the MERAXES semi-analytic model (Mutch et al., 2016) is utilised. This chapter explains the relevant galaxy formation physics implemented in the model. A full description of the model can be found in Mutch et al. (2016) and Qin et al. (2017).

2.1 Gas infall

At each time step, the MERAXES model walks forward the input halo merger trees, adn the first step is to determine the infall gas accreted to each friends-of-friends halo. The mass of infall gas is computed by

$$m_{\rm infall} = f_{\rm mod} f_{\rm b} M_{\rm vir} - \sum m^i_* + m^i_{\rm cold} + m^i_{\rm hot} + m^i_{\rm ejected}, \tag{2.1}$$

where $f_{\rm b}$ is the global baryon fraction, $f_{\rm mod}$ is the baryon modifier, and $M_{\rm vir}$ is the halo mass. The baryon modifier is related to the reionisation feedback described in Section 2.6. The baryonic components of galaxies are labelled as m^i_*, $m^i_{\rm cold}$, $m^i_{\rm hot}$, and $m^i_{\rm ejected}$, which will be introduced in the following sections. The summation is over all galaxies in the friends-of-friends group.

2.2 Cooling

The infall gas is put into the hot gas component and is assumed to form a quasi-static isothermal sphere, with density profile:

$$\rho_{\rm hot}(r) = \frac{m_{\rm hot}}{4\pi R_{\rm vir} r^2}, \tag{2.2}$$

where $R_{\rm vir}$ is the virial radius. The temperature of the isothermal sphere is assumed to the virial temperature of the host halo:

$$T_{\rm vir} = 35.9 \left(\frac{V_{\rm vir}}{\rm km/s}\right)^2 {\rm K}, \tag{2.3}$$

where $V_{\rm vir}$ is the virial velocity. The cooling rate at a given radius can be computed by

$$t_{\rm cool}(r) = \frac{3}{2}\frac{\bar{\mu} m_{\rm p} k_{\rm B} T}{\rho_{\rm hot}(r)\Lambda(T, Z)}, \tag{2.4}$$

where $\bar{\mu} m_{\rm p}$ is the mean gas particle mass, Z is the metallicity, and $\Lambda(T, Z)$ is the cooling function given by Sutherland & Dopita (1993). When gas cools down, it flow into the central region and becomes fuel of star formation. This process is modelled as mass transfer between the hot and cold gas components of galaxies. Define the cooling radius $r_{\rm cool}$ as the radius at which $t_{\rm cool} = t_{\rm dyn} = R_{\rm vir}/V_{\rm vir}$. There are three cooling regimes depending on whether a hydrostatic equilibrium can be reached and the atomic cooling threshold. The transfer rate can be calculated by

$$\dot{m}_{\rm cool} = \begin{cases} \frac{m_{\rm hot}}{t_{\rm dyn}}, & r_{\rm cool} \geq R_{\rm vir}, T \geq 10^4{\rm K}, \\ \frac{r_{\rm cool}}{R_{\rm vir}}\frac{m_{\rm hot}}{t_{\rm dyn}}, & r_{\rm cool} < R_{\rm vir}, T \geq 10^4{\rm K}, \\ 0, & T < 10^4{\rm K}. \end{cases} \tag{2.5}$$

2.3 Star formation

In the MERAXES model, cold gas is assumed to form a cold disk in the central region of the host halo. Using a stability argument, Kauffmann (1996) suggested that star

formation only occurs if the mass of the disk is greater than the critical mass

$$m_{\rm crit} = \Sigma_{\rm SF} \left(\frac{V_{\rm max}}{100{\rm km/s}} \right) \left(\frac{r_{\rm disk}}{10{\rm kpc}} \right) \times 10^{10}\,{\rm M}_{\odot}, \tag{2.6}$$

with

$$r_{\rm disk} = 3R_{\rm vir} \frac{\lambda}{\sqrt{2}}, \tag{2.7}$$

where $V_{\rm max}$ and λ are the maximum circular velocity and spin parameter (defined by the Bullock et al. (2001)) of the host halo. The star formation rate can be calculated by

$$\dot{m}_{\rm star} = \alpha_{\rm SF} \frac{m_{\rm cold} - m_{\rm crit}}{t_{\rm dyn,disk}}, \tag{2.8}$$

where $t_{\rm dyn,\ disk} = r_{\rm disk}/V_{\rm max}$ is the dynamical time of the disk. This star formation model contains two free parameters, i.e. $\Sigma_{\rm SF}$ and $\alpha_{\rm SF}$, which can be calibrated against observations.

2.4 Supernova feedback

Supernova feedback plays an important role in regulating star formation. In semi-analytic models, this process is represented by mass transfer from the cold gas component to the hot gas and ejected gas components. The supernova feedback model implemented in MERAXS is developed from Guo et al. (2011), and relaxes the instantaneous recycling approximation by taking into account the supernova events from a short star formation history.

Mass transfer from cold gas to hot gas can be computed as

$$\Delta m_{\rm reheat} = \begin{cases} \eta \Delta m_{\rm new}, & \Delta E_{\rm SN} \geq \Delta E_{\rm hot} \\ \frac{\Delta E_{\rm SN}}{1/2 V_{\rm vir}^2}, & \Delta E_{\rm SN} < \Delta E_{\rm hot} \end{cases}, \tag{2.9}$$

with

$$\Delta E_{\rm hot} = \frac{1}{2} \eta \Delta m_{\rm new} V_{\rm vir}^2, \tag{2.10}$$

where η is the mass loading factor, $\Delta m_{\rm new}$ is the mass of new formed stars, $\Delta E_{\rm SN}$ is the supernova energy injected into the interstellar medium (ISM), and $\Delta E_{\rm host}$ is the energy increase of the hot gas halo due to the mass transfer. The above equations state that

if $E_{\rm SN} \geq \Delta E_{\rm hot}$, the reheated mass can determined using the mass loading argument, while $E_{\rm SN} < \Delta E_{\rm hot}$, the reheated mass is limited by the supernova energy. Furthermore, in the case where $E_{\rm SN} \geq \Delta E_{\rm hot}$, the hot gas mass should be decreased due to energy conservation. The additional mass is transferred to the ejected gas component using

$$\Delta m_{\rm eject} = \frac{\Delta E_{\rm SN} - \Delta E_{\rm hot}}{1/2 V_{\rm vir}^2}. \tag{2.11}$$

Let the time step of the simulation be Δt, the supernovae energy injected into the ISM during this period can be calculated by

$$\Delta E_{\rm SN} = \epsilon \times \int_t^{t+\Delta t} dt' \int_0^{\infty} d\tau \, \frac{d\varepsilon}{d\tau} \psi(t' - \tau), \tag{2.12}$$

where $\psi(t' - \tau)$ is the star formation history and $(d\varepsilon/d\tau)d\tau$ is the energy released by type-II supernova from stars with age between τ to $\tau + d\tau$. Similarly, $\Delta m_{\rm new}$ can be calculated by

$$\Delta m_{\rm new} = \frac{\int_t^{t+\Delta t} dt' \int_0^{\infty} d\tau \, \frac{d\varepsilon}{d\tau} \psi(t' - \tau)}{\int_0^{\infty} d\tau \, \frac{d\varepsilon}{d\tau}}. \tag{2.13}$$

Mutch et al. (2016) estimates $d\varepsilon/d\tau$ using the stellar life time obtained by Portinari et al. (1998) and the Salpeter (1955) initial mass function (IMF). A different approach will be introduced in Section 4.2.

In the supernova feedback described above, the mass loading factor η and the energy coupling efficiency ϵ are based on empirical relations, which contain several free parameters. Following Guo et al. (2011), Mutch et al. (2016) adopt

$$\eta = \alpha_{\rm mass} \left[0.5 + \left(\frac{V_{\rm max}}{V_{\rm mass}} \right)^{-\beta_{\rm mass}} \right], \tag{2.14}$$

$$\epsilon = \alpha_{\rm energy} \left[0.5 + \left(\frac{V_{\rm max}}{V_{\rm energy}} \right)^{-\beta_{\rm energy}} \right], \tag{2.15}$$

where $\alpha_{\rm mass}$, $\beta_{\rm mass}$, $V_{\rm mass}$, $\alpha_{\rm energy}$, $\beta_{\rm energy}$, and $V_{\rm energy}$ are all free parameters. Mutch et al. (2016) employ a maximum mass loading factor $\eta_{\rm mass}^{\rm max}$, which is also a free parameter. The maximum of the energy coupling efficiency is set to be unity due to energy conservation.

2.5 Mass recycling and metal enrichment

Supernovae explosion releases materials into the ISM, which then become fuel of star formation. Metals are also produced by this process. The mass of released materials can be computed by

$$\Delta m_{\rm recycle} = \int_t^{t+\Delta t} dt' \int_0^{\infty} d\tau \, \frac{dy}{d\tau} \psi(t'-\tau), \tag{2.16}$$

where $(dy/d\tau)d\tau$ is the mass produced by type-II supernova from stars with age τ to $\tau + d\tau$. This quantity is also known as the yield, and varies with different elements. Mutch et al. (2016) estimates the yield using a similar approach of calculating $d\varepsilon/d\tau$. An improved approach will be discussed in 4.2. The mass of recycled materials is added to the cold gas mass.

2.6 Reionisation feedback

To implement self-consistent reionisation feedback, MERAXES is coupled with the 21CMFAST semi-numerical reionisation model. The 21CMFAST model applies a excursion set algorithm to probe the ionising structure of the inter-galactic media (IGM). Practically, we divide the simulation box to a cubic grid. At each time step, after evolving galaxy properties, MERAXES provides a star formation rate grid to 21CMFAST. Then, 21CMFAST estimates neutral fraction $x_{\rm H}$ and ultraviolet background (UVB) intensity grids J_{21} for the calculation of reionisation feedback. The UVB can reduce the amount of gas in a halo through photoevaporation. Using 1D hydrodynamical simulation, Sobacchi & Mesinger (2013) provided an expression of the baryon modifier with the impact of UVB:

$$f_{\rm mod} = 2^{-M_{\rm filt}/M_{\rm vir}}, \tag{2.17}$$

with

$$M_{\rm filt} = M_0 J_{21}^a \left(\frac{1+z}{10}\right)^b \left[1 - \left(\frac{1+z}{1+z_{\rm ion}}\right)^c\right]^d, \tag{2.18}$$

where $z_{\rm ion}$ is the redshift at which the halo was first exposed to the UVB. Sobacchi & Mesinger (2013) suggested that $M_0 = 2.8 \times 10^9 M_\odot$, $a = 0.17$, $b = -2.1$, $c = 2.0$, and

$d = 2.5$. The baryon modifier given by Equation 2.17 is coupled with MERAXES via Equation 2.1, which computes the amount of infall gas.

Chapter 3

Clustering segregation with stellar mass and UV-luminosity

3.1 Introduction

The first application of the MERAXES semi-analytic model presented in the thesis is about galaxy clustering, which provides a probe of the host halo mass of galaxies. The clustering strength is commonly described by the two-point correlation function, which measures the probability of finding galaxy pairs at given spatial separations. Mo & White (1996) used the extended Press-Schechter formalism (Bond et al., 1991) to show that the ratio between the correlation functions of halos and the underlying matter depends on halo mass. This ratio is known as bias. Since galaxies reside in halos, the bias links galaxy clustering to the mass of their host halos (see Cooray & Sheth, 2002 for a review).

The dependence of clustering strength on galaxy properties is known as clustering segregation, and reveals the correlation between galaxy properties and halo mass. At high redshifts, clustering segregation is observed for Lyman-break galaxies (LBGs) with UV-luminosity (Lee et al., 2006; Hildebrandt et al., 2009; Barone-Nugent et al., 2014; Harikane et al., 2016, 2018) and stellar mass (Ishikawa et al., 2017; Durkalec et al., 2018). One basic conclusion from these studies is that more luminous and larger stellar mass galaxies are more clustered, and therefore reside in more massive halos.

In the context of hierarchical galaxy formation, this correlation between halo and galaxy properties is unsurprising, since halo mass is closely related to the gas reservoir available for star formation and those processes which impact on it. For instance, in the low mass regime, supernova (SN) feedback can effectively suppress star formation (Wyithe & Loeb, 2013; Hopkins et al., 2014; Duffy et al., 2014). Therefore, it is of particular interest to explore which galaxy property is more tightly correlated with the host halo mass. While UV-luminosity is directly related to the current star formation rate, stellar mass provides integrated information over the star formation history. For this reason, it is expected that, when splitting the same sample by luminosity and stellar mass, clustering segregation with stellar mass should be larger than with UV magnitude.

This work is divided into two parts. Firstly, I measure clustering segregation with both stellar mass and UV-luminosity at $z \sim 4-7$ using data from the Hubble Space Telescope (HST) (Section 3.2). Secondly, I compare the measurements with predictions from MERAXES (Section 3.3). Finally, Section 3.4 summarises this work.

3.2 Measuring clustering segregation

3.2.1 Observational data

The galaxy catalogue used in the clustering measurements of this work is provided by my collaborator, Pascal Oesch. It is based on the photometric catalogue from Bouwens et al. (2015), who selected Lyman break galaxies (LBGs) at $z \sim 4-7$ based on the *Hubble Space Telescope* (HST) data in all the CANDELS fields, as well as the very deep XDF and HUDF09 parallel fields. In particular, the galaxy sample is drawn from the *XDF* (Illingworth et al., 2013), HUDF-091 and HUDF-092 (Bouwens et al., 2011), *CANDELS-GN* and *CANDELS-GS* (Grogin et al., 2011; Koekemoer et al., 2011), *ERS* (Windhorst et al., 2011), and *CANDELS-UDS*, *CANDELS-COSMOS* and *CANDELS-EGS* (Grogin et al., 2011; Koekemoer et al., 2011). These survey regions span an aggregate of $\sim 700\,\mathrm{arcmin}^2$ in the sky, and $\sim 10,000$ LBGs are identified. Photometric redshifts of these sources are estimated using the EAZY code (Brammer et al., 2008). For more information on the LBG selection and the photometric redshifts see Bouwens et al. (2015).

Pascal combined the HST photometry with the large archive of *Spitzer/IRAC* legacy data available in the CANDELS fields (Ashby et al., 2013, 2015), which includes the ultra-deep IGOODS/IUDF and GREATS surveys (Labbé et al., 2015, Labbé et al. 2018, in prep.) in the GOODS fields. IRAC photometry is measured in circular apertures after subtracting the contaminating flux of neighboring galaxies using the code *mophongo* (Labbé et al., 2006, Labbé et al., 2018, in prep.), which is similar to the code TPHOT (Merlin et al., 2016).

Pascal measured stellar masses of galaxies based on SED fitting to the HST+Spitzer photometry using ZEBRA+ (Oesch et al., 2010). The synthetic template set used here is based on Bruzual & Charlot (2003) with a constant star-formation history, sub-solar metallicities (0.2 $Z_\odot$) and a Chabrier (2003) initial mass function (IMF). Nebular continuum and emission lines are added self-consistently based on the number of ionizing photons emitted by each SED and assuming line ratios relative to Hβ as tabulated by Anders & Fritze-v. Alvensleben (2003). Dust extinction is applied using the attenuation curve by Calzetti et al. (2000).

Following Bouwens et al. (2015), the absolute magnitudes, $M_{\rm UV}$, are computed based on the fluxes in the photometric band that is closest to rest-frame 1600Å. I first fit the $M_\star - M_{\rm UV}$ relation for the LBG sample, which will be used in the clustering analysis to compare stellar mass and luminosity segregation. The form of the relation is assumed to be

$$\log_{10} M_\star^{\rm Fit} = \frac{d\log_{10} M_\star}{dM_{\rm UV}} (M_{\rm UV} + 19.5) + \log_{10} M_{\star(M_{\rm UV}=-19.5)}, \tag{3.1}$$

where mass is in units of $M_\odot$. The log-likelihood is then constructed as

$$\ln \mathcal{L} = -\frac{1}{2} \sum_i \left[\frac{\log_{10}(M_{\star i}^{\rm Obs} / M_\star^{\rm Fit})^2}{\Delta^2} + \ln(2\pi\Delta^2) \right], \tag{3.2}$$

where the sum is over all LBGs, and Δ is a mass-independent free parameter representing scatter in the $M_\star - M_{\rm UV}$ relation. I adopt a Bayesian approach to perform the fit, assume constant priors for all parameters, and apply the EMCEE MCMC sampler developed by Foreman-Mackey et al. (2013). The resulting $M_\star - M_{\rm UV}$ relations are shown in Figure 3.1. Best-fit parameters are given in Table 3.1. I find that the $M_\star - M_{\rm UV}$ relations are close to linear ($M_\star \propto L$) and that the scatter in stellar mass at fixed luminosity is

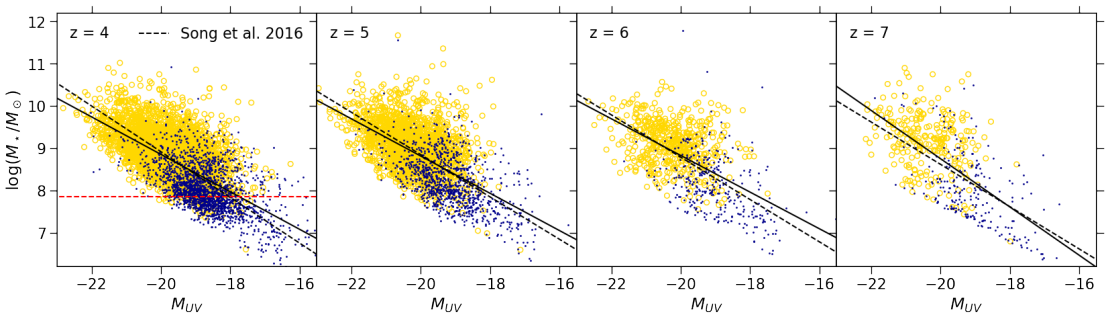


FIGURE 3.1: Scatter plots of the $M_\star - M_{\rm UV}$ relations. The absolute magnitude $M_{\rm UV}$ is based on the flux in the filter whose centre wavelength is closest to rest-frame 1600Å and the stellar mass is measured by SED fitting including Spitzer data. For yellow empty circles, the sources have at least one Spitzer band ($3.6\mu m$ and $4.5\mu m$) with S/N> 3, while small blue dots show the remaining sources. Black solid lines are best-fit results. The corresponding parameters are summarised in Table 3.1. For comparison, black dashed lines show the results from Song et al. (2016), which are converted to a Chabrier (2003) IMF by subtracting 0.24 dex in the stellar mass. The dashed horizontal line shows the lower bound of the least massive stellar mass bin for the clustering measurement at $z \sim 4$.

~ 0.5 dex. Even though the best-fit slopes are slightly shallower, my measurements are consistent with the recent study from Song et al. (2016).

In this work, every galaxy in our HST sample has an estimate of stellar mass irrespective of the quality of Spitzer data. Low S/N ratios of Spitzer bands could make the stellar masses less precise. To investigate this, in Figure 3.1, galaxies that have at least one Spitzer band ($3.6\mu m$ and $4.5\mu m$) with S/N> 3 are shown as yellow empty circles in Figure 3.1, while the others are shown as small blue dots. No systematic offset is found between them. Since the sample is large enough at $z \sim 4$, I use multiple stellar mass and luminosity bins for the clustering measurements, and avoid using bins that have no lower bound to reduce possible effects due to low S/N ratios of Spitzer data. The lower bound for the least massive stellar mass bin is shown as red dashed line in the corresponding panel of Figure 3.1. At all other redshifts, limited by the sample size, I include all galaxies and use two bins to examine clustering segregation with stellar mass and luminosity. The fraction of LBGs that are included and have at least one Spitzer band with S/N> 3 is 74%, 63%, 53%, and 47% at $z \sim 4, 5, 6$, and 7 respectively. The advantage of this approach is that the completeness of sample LBGs, which is defined by the selection, is not affected by Spitzer data, and it is important to use the same set of galaxies to compare the clustering segregation between UV-luminosity and stellar mass. I will discuss possible effects on clustering measurements due to uncertainties in stellar mass in Section 3.3.2.

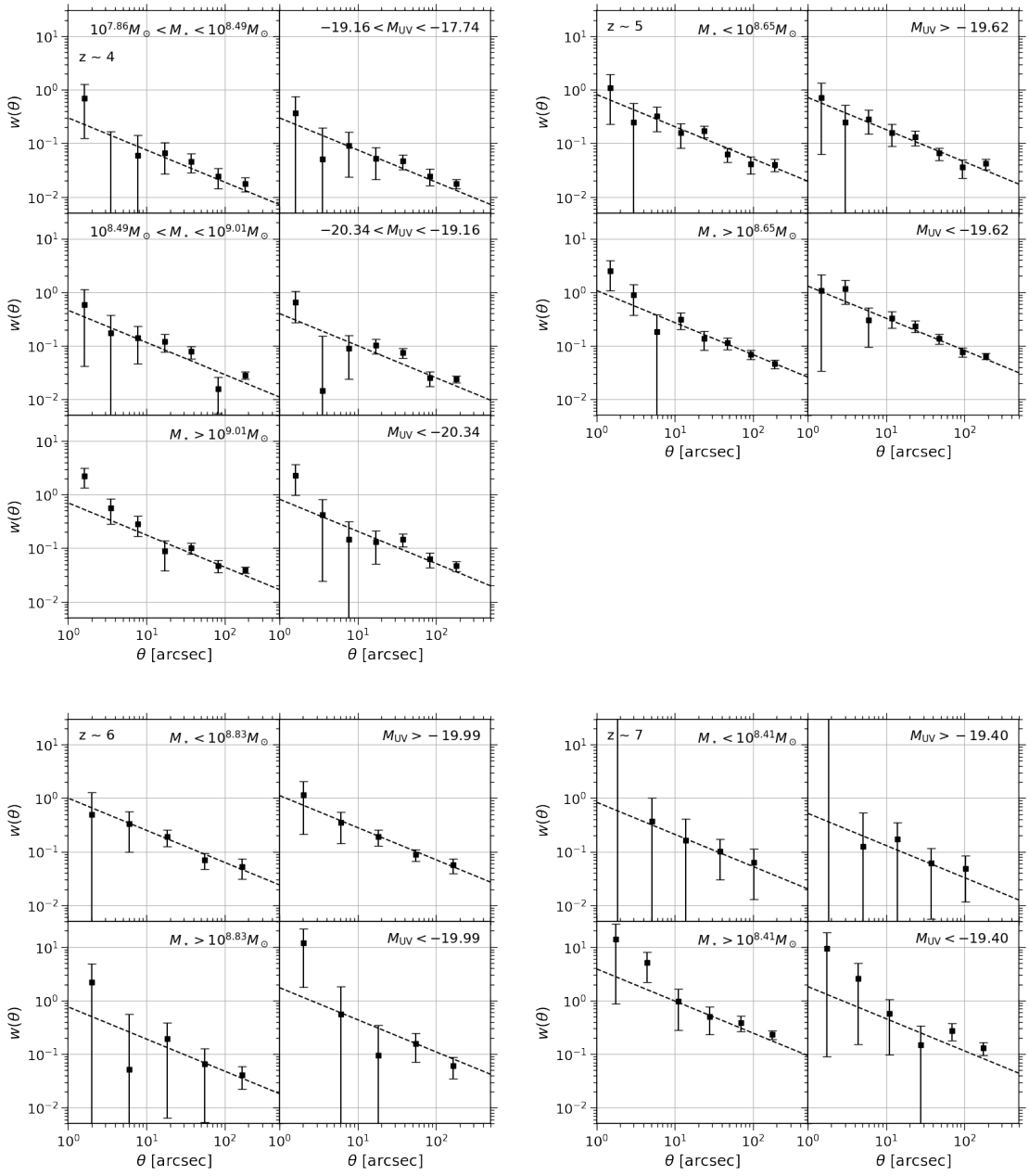


FIGURE 3.2: Measured ACFs and their best-fit power laws $A_\omega(\theta/1'')^{-0.6}$. Top left, top right, bottom left, and bottom right panels show the measurements at $\bar{z} = 3.8$, 5.0, 5.9, and 6.8 respectively. For each panel, the first and second columns illustrate the stellar mass and luminosity subsamples respectively. In all plots, black squares with error bars are measured ACFs, which are averaged over all fields using inverse-variance weighting, and dashed lines are best-fit power laws.

TABLE 3.1: Best-fit parameters of the $M_\star - M_{\rm UV}$ relations

$\bar{z}$	$d\log_{10} M_\star/dM_{\rm UV}$	$\log_{10} M_{\star(M_{\rm UV}=-19.5)}$	Δ
3.8	$-0.44^{+0.01}_{-0.01}$	$8.64^{+0.01}_{-0.01}$	$0.501^{+0.005}_{-0.005}$
5.0	$-0.44^{+0.01}_{-0.01}$	$8.60^{+0.01}_{-0.01}$	$0.555^{+0.008}_{-0.007}$
5.9	$-0.43^{+0.02}_{-0.02}$	$8.61^{+0.02}_{-0.02}$	$0.621^{+0.017}_{-0.013}$
6.8	$-0.57^{+0.03}_{-0.03}$	$8.47^{+0.03}_{-0.03}$	$0.734^{+0.028}_{-0.020}$

Notes. - Mass unit is $M_\odot$. Quoted errors are 16% and 84% percentiles of the marginalised distributions estimated by the MCMC sampler.

3.2.2 Estimating the angular correlation function

My approach follows Barone-Nugent et al. (2014). I start by determining the angular correlation functions (ACFs), which measure the excess probability of finding galaxy pairs with angular separations between θ and $\theta + \delta\theta$. The ACF estimator proposed by Landy & Szalay (1993) is applied, i.e.

$$\omega_{\rm obs}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \tag{3.3}$$

where $DD(\theta)$, $DR(\theta)$ and $RR(\theta)$ are the probability of finding galaxy-galaxy, galaxy-random and random-random pairs respectively. These probabilities are calculated by counting all pairs at separations between θ to $\theta + \delta\theta$, and normalising by the total number of pairs. Estimates of $DR(\theta)$ and $RR(\theta)$ require a catalogue of uniformly-distributed random points. This is generated by a random Poisson process. For each field, the random catalogue contains 10,000 points uniformly placed within the corresponding survey regions. I measure the ACFs in logarithmic bins, and estimate errors by bootstrap resampling (Ling et al., 1986). I construct bootstrap subsamples by replacing individual galaxies and perform the resampling for $\sim$ 500 times. This approach is also used in Barone-Nugent et al. (2014) and Harikane et al. (2016). The latter also found that neglecting off-diagonal elements of the covariance matrix does not have significant impacts on the results. Therefore, off-diagonal terms are not considered in this work. In order to investigate the clustering dependence on both stellar mass and UV magnitude, the total sample is split into subsamples. I choose bins such that they satisfy the $M_\star - M_{\rm UV}$ relation at each redshift. The bin cuts are listed in Table 3.2.

Since the area of each survey region is finite, the observed ACFs are affected by border effects. This is corrected by an additive constant, which is known as the intergral

constrain (IC). Following Roche & Eales (1999), we have

$$\omega_{\rm true}(\theta) = \omega_{\rm obs}(\theta) + {\rm IC}, \tag{3.4}$$

with

$${\rm IC} = \frac{1}{\Omega^2}\iint \omega_{\rm true}(\theta)\, d\Omega_1 d\Omega_2 = \frac{\sum_i RR(\theta_i)\omega_{\rm true}(\theta_i)}{\sum_i RR(\theta_i)}, \tag{3.5}$$

where $\omega_{\rm true}(\theta)$ is the fitting model of the ACF.

I assume the ACFs to be a power law

$$\omega_{\rm true}(\theta) = A_\omega \left(\frac{\theta}{1''}\right)^{-\beta}, \tag{3.6}$$

and construct the log-likelihood using

$$\ln \mathcal{L} = -\frac{1}{2}\sum_{\rm fields}\sum_i \left[\frac{\omega_{\rm obs}(\theta_i) - A_\omega(\theta_i^{-\beta} - {\rm IC}/A_\omega)}{\sigma(\theta_i)}\right]^2, \tag{3.7}$$

where sums are over all bins i and over all survey fields. This is equivalent to measuring the average ACF of all fields using inverse-variance weighting. Since this approach requires an estimate of the ACF in each individual field, I only include fields that are deep enough such that the mean separation of galaxy pairs is smaller than 100 arcsec. The number of LBGs that enter into the analysis for each subsample is given in Table 3.2. Moreover, the dependence of the IC on the fitting model results in some degeneracy between A_ω and β (Lee et al., 2006), I therefore fix $\beta = 0.6$ following Lee et al. (2006) and Barone-Nugent et al. (2014). In addition, since the area of XDF, HUDF-091, and HUDF-092 is only 4.7 arcmin^2 (i.e. one WFC3/IR pointing), the counted number of galaxy pairs in these fields decreases when the angular separation is greater than ~ 140 arcsec. I therefore only include separations smaller than that in the likelihood function.

The amplitude of the ACF A_ω could be weakened by contamination of lower redshift sources. I reduce this effect by removing all LBGs whose best-fit photometric redshift indicates that it might be a low redshift contaminant ($z_{\rm phot} < 2$). It is also noted that Harikane et al. (2016) used the contamination fraction estimated by Bouwens et al. (2015) to correct for this effect. They found that the difference is insignificant compared with the statistical error. Thus, no further treatment is employed to correct the effect of contamination.

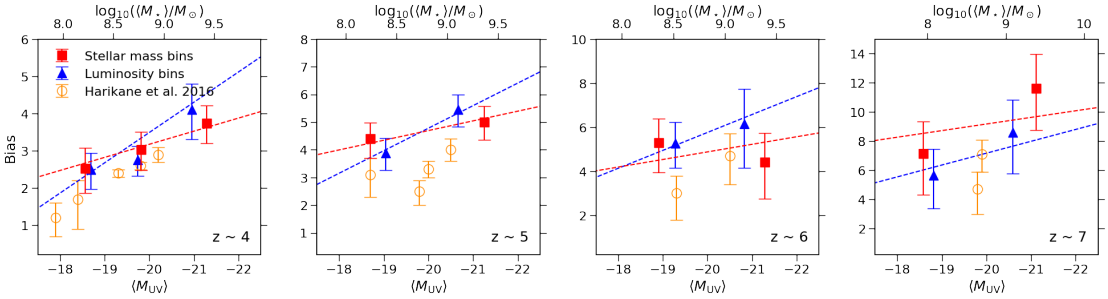


FIGURE 3.3: Measured galaxy biases as a function of mean stellar mass and mean UV flux. The bias is computed by equation 3.11 and 3.12. For all panels, squares (triangles) are corresponding to stellar mass (luminosity) split samples, and are plotted against the top (bottom) axis. Dashed lines are best-fit linear relations of the biases, which are the same with those in Figure 3.4. Scales of these double x-axes are such chosen that they satisfy the $M_\star - M_{\rm UV}$ relation at each redshift. Orange circles with error bars are biases estimated by Harikane et al. (2016) as a function of mean UV magnitude, which are listed in their Table 5.

3.2.3 Estimating the correlation length and bias

Real space parameters are obtained by applying the Limber transform to the ACFs. The real-space correlation $\xi(r)$ provides three dimensional information on galaxy clustering, which is also approximated by a power law,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{3.8}$$

where r_0 is called the correlation length. In this case, the Limber transform takes the form (Peebles, 1980)

$$\beta = \gamma - 1, \tag{3.9}$$

and

$$A_\omega = r_0^\gamma B\left(\frac{1}{2}, \frac{\gamma-1}{2}\right) \frac{\int_0^\infty dz\, N^2(z) {d_{\rm H}}^{-1} d_{\rm A}^{1-\gamma} (1+z)^{1-\gamma}}{\left[\int_0^\infty dz\, N(z)\right]^2}, \tag{3.10}$$

with

$$d_{\rm H} = \frac{c}{H(z)}, \quad d_{\rm A} = \frac{1}{1+z}\int_0^\infty d_{\rm H}\, dz,$$

where $B(x, y)$ is the beta function, $N(z)$ is the redshift distribution function of sample galaxies, and $H(z)$ is the Hubble parameter as a function of redshift. The above equations link A_ω and β to the power law parameters in the real space. $N(z)$ is estimated using the photometric redshifts of each LBG.

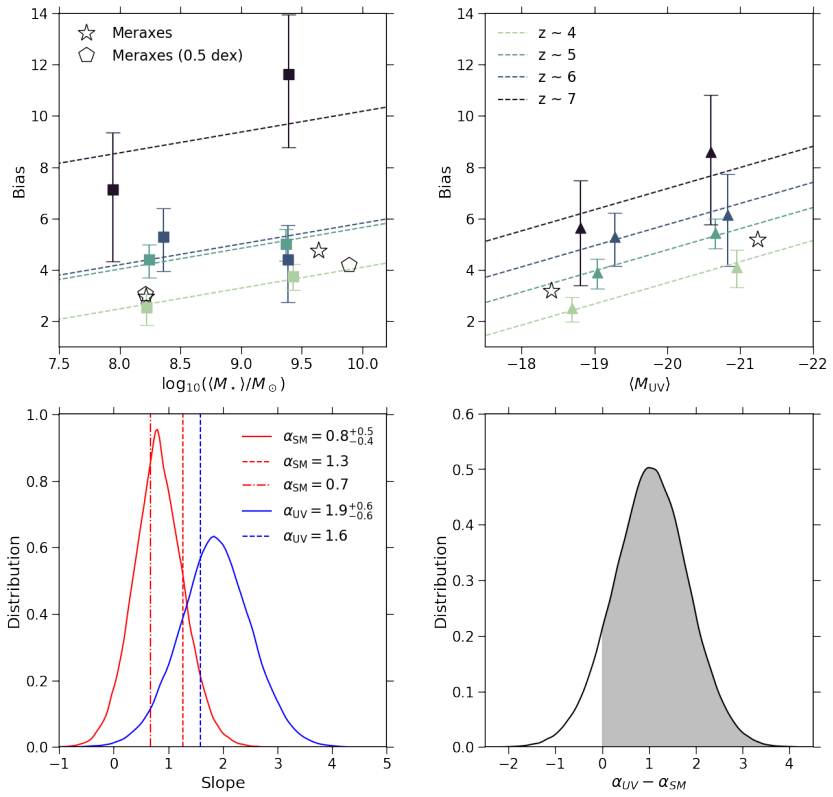


FIGURE 3.4: Results of a straight line fit of measured biases over all redshifts (assuming the same slope). In top panels, dashed lines are the best-fit results. Squares and triangles with error bars are the data that is used for the fits. Empty stars and pentagons are predicted biases from the semi-analytic model MERAXES. When computing those pentagons, 0.5 dex Gaussian scatters are added to the stellar mass of each model galaxy. In the bottom left panel, solid lines depict the marginalised distributions of the slope. I show the medians and $1-\sigma$ percentiles of the distributions in the top right corner. Dashed vertical lines show the slopes derived from the model, and the dot dashed line gives the result in the case where the model stellar mass has scatter added. The values of these vertical lines are also shown in the top right corner. In this panel, all red and blue lines correspond to stellar mass and luminosity split samples respectively. Bottom right panel shows the distribution of $\alpha_{\rm UV}-\alpha_{\rm SM}$ obtained by subtracting the samples of $\alpha_{\rm UV}$ and $\alpha_{\rm SM}$. The area of the shaded region is $\simeq 90\%$, which is the probability that $\alpha_{\rm UV} > \alpha_{\rm SM}$.

I derive the bias using the ratio between the variance of the galaxy and the matter correlation functions smoothed by a top-hat with radius $8h^{-1}$Mpc:

$$b = \frac{\sigma_{8,\ \rm g}}{\sigma_8}, \tag{3.11}$$

where (Peebles, 1980)

$$\sigma^2_{8,\ \rm g} = \frac{72(r_0/8h^{-1}{\rm Mpc})^\gamma}{(3-\gamma)(4-\gamma)(6-\gamma)2^\gamma}. \tag{3.12}$$

TABLE 3.2: Summary of clustering measurements at $z \sim 4-7$.

$\bar{z}$ (1)	Sample (2)	Cut (3)	Number (4)	$\overline{M}_\star$ (5)	$\overline{M}_{\rm UV}$ (6)	A_ω (7)	r_0 (8)	b (9)
3.8	Stellar mass	$10^{7.86}M_\odot < M_* < 10^{8.49}M_\odot$	1636	$10^{8.22}$	-19.28	$0.30^{+0.18}_{-0.16}$	$2.7^{+0.8}_{-0.9}$	$2.5^{+0.5}_{-0.7}$
		$10^{8.49}M_\odot < M_* < 10^{9.01}M_\odot$	1691	$10^{8.77}$	-19.82	$0.46^{+0.18}_{-0.18}$	$3.4^{+0.7}_{-0.8}$	$3.0^{+0.5}_{-0.6}$
		$M_* > 10^{9.01}M_\odot$	1514	$10^{9.43}$	-20.44	$0.70^{+0.20}_{-0.20}$	$4.4^{+0.7}_{-0.8}$	$3.7^{+0.5}_{-0.5}$
	Luminosity	$-19.16 < M_{\rm UV} < -17.74$	2012	$10^{8.71}$	-18.70	$0.30^{+0.12}_{-0.14}$	$2.8^{+0.6}_{-0.7}$	$2.5^{+0.4}_{-0.5}$
		$-20.34 < M_{\rm UV} < -19.16$	2460	$10^{9.06}$	-19.74	$0.40^{+0.12}_{-0.12}$	$3.0^{+0.5}_{-0.6}$	$2.8^{+0.4}_{-0.4}$
		$M_{\rm UV} < -20.34$	957	$10^{9.52}$	-20.95	$0.82^{+0.32}_{-0.32}$	$4.9^{+1.0}_{-1.2}$	$4.1^{+0.7}_{-0.8}$
5.0	Stellar mass	$M_* < 10^{8.65}M_\odot$	1191	$10^{8.24}$	-19.62	$0.82^{+0.26}_{-0.26}$	$4.3^{+0.7}_{-0.8}$	$4.4^{+0.6}_{-0.7}$
		$M_* > 10^{8.65}M_\odot$	1554	$10^{9.37}$	-20.56	$1.08^{+0.26}_{-0.28}$	$4.9^{+0.7}_{-0.8}$	$5.0^{+0.6}_{-0.7}$
	Luminosity	$M_{\rm UV} > -19.62$	1007	$10^{9.04}$	-19.04	$0.72^{+0.22}_{-0.24}$	$3.7^{+0.7}_{-0.7}$	$3.9^{+0.5}_{-0.6}$
		$M_{\rm UV} < -19.62$	1661	$10^{9.47}$	-20.66	$1.30^{+0.28}_{-0.30}$	$5.4^{+0.7}_{-0.7}$	$5.4^{+0.5}_{-0.6}$
5.9	Stellar mass	$M_* < 10^{8.83}M_\odot$	342	$10^{8.36}$	-19.80	$1.00^{+0.50}_{-0.50}$	$4.5^{+1.2}_{-1.4}$	$5.3^{+1.1}_{-1.3}$
		$M_* > 10^{8.83}M_\odot$	293	$10^{9.39}$	-20.46	$0.76^{+0.66}_{-0.66}$	$3.5^{+1.4}_{-1.6}$	$4.4^{+1.3}_{-1.7}$
	Luminosity	$M_{\rm UV} > -19.99$	450	$10^{9.38}$	-19.28	$1.12^{+0.48}_{-0.46}$	$4.5^{+1.0}_{-1.2}$	$5.3^{+0.9}_{-1.1}$
		$M_{\rm UV} < -19.99$	205	$10^{9.43}$	-20.83	$1.74^{+1.16}_{-1.14}$	$5.3^{+1.8}_{-2.1}$	$6.1^{+1.6}_{-2.0}$
6.8	Stellar mass	$M_* < 10^{8.41}M_\odot$	118	$10^{7.94}$	-19.62	$0.84^{+0.80}_{-0.80}$	$5.4^{+2.3}_{-2.6}$	$7.1^{+2.2}_{-2.8}$
		$M_* > 10^{8.41}M_\odot$	149	$10^{9.39}$	-20.53	$3.90^{+1.84}_{-1.84}$	$10.3^{+2.7}_{-3.1}$	$11.6^{+2.3}_{-2.8}$
	Luminosity	$M_{\rm UV} > -19.40$	130	$10^{9.15}$	-18.81	$0.52^{+0.54}_{-0.52}$	$4.1^{+1.8}_{-2.1}$	$5.7^{+1.8}_{-2.3}$
		$M_{\rm UV} < -19.40$	200	$10^{9.56}$	-20.60	$1.82^{+1.22}_{-1.22}$	$6.9^{+2.3}_{-2.8}$	$8.6^{+2.2}_{-2.8}$

Notes. - (1) Mean redshift. (2) Sample name. (3) Stellar mass and luminosity cuts. The unit of stellar mass cut is $M_\odot$. (4) Number of galaxies in the subsample. (5) Mean stellar mass in a unit of $M_\odot$. (6) Mean UV fluxes. (7) Amplitude of ACF with fixed $\beta = 0.6$. (8) Correlation length in a unit of h^{-1}Mpc. (9) Bias.

3.2.4 Results and discussion

Figure 3.2 shows measured ACFs and best-fit power laws for both stellar mass and luminosity subsamples. Results for the angular correlation function amplitude (A_ω), the correlation length (r_0), and bias (b) are summarised in Table 3.2. All quantities given in the table are the most probable value and the corresponding 1σ error.

Clustering segregation is observed with both stellar mass and luminosity. From Figure 3.2, it is clear that the ACFs increase with luminosity at all redshifts. Similar trends are also observed for different stellar mass subsamples. However, the segregation with stellar mass is found to be weaker than with luminosity. To further compare the clustering dependence of these two properties, I plot the measured bias as a function of mean stellar mass and flux in Figure 3.3. The bias increases with stellar mass, from $b = 2.7^{+0.5}_{-0.6}$ to $b = 3.6^{+0.5}_{-0.5}$ at $z \sim 4$, and $b = 4.2^{+0.6}_{-0.7}$ to $4.8^{+0.6}_{-0.7}$ at $z \sim 5$. The segregation of bias with luminosity is more obvious. At $z \sim 4$, the bias for the brightest sample is $b = 4.1^{+0.7}_{-0.8}$, while that of the faintest sample is smaller at $b = 2.5^{+0.4}_{-0.5}$. Similarly, at $z \sim 5$, the bias has an increase between the faint and bright samples, from $b = 3.8^{+0.5}_{-0.6}$ to $b = 5.3^{+0.6}_{-0.6}$. At

$z \gtrsim 6$, the LBG sample is smaller, and uncertainties become much larger. While there is a small increase of the bias with mean UV flux at $\bar{z} = 5.9$, no significant segregation with stellar mass is detected. For samples at $z \sim 7$, I find that the bias increases from $b = 7.4^{+2.0}_{-2.6}$ to $b = 11.7^{+2.2}_{-2.7}$, and from $b = 5.6^{+1.8}_{-2.2}$ to $b = 8.6^{+2.2}_{-2.8}$, for stellar mass and luminosity bins respectively. As a comparison, I also plot the bias estimated by Harikane et al. (2016) from their power law fits as a function of mean UV magnitude. I find that the trends of clustering dependence on luminosity are consistent, while the offsets on biases themselves could be due to different methods of computing them. To summarise, I find that both more massive and more luminous galaxies are more highly clustered, implying that they are hosted by more massive dark matter halos.

On the other hand, the comparison of segregation between stellar mass and luminosity disagrees with my prior expectations, especially at $\bar{z} = 3.8$ and 5.0. In particular, I find that the clustering dependence is larger for luminosity than for stellar mass. In order to quantify this trend, I fit the measured biases of the lightest (faintest) and heaviest (brightest) by straight lines, i.e.

$$b = \alpha_{\rm SM} \log_{10}(M_\star/M_\odot) + \text{const}, \tag{3.13}$$

and

$$b = 1.1\, \alpha_{\rm UV} \log_{10}(L_{\rm UV}/L_\odot) + \text{const}. \tag{3.14}$$

If clustering segregation with both properties were the same, one would expect that the ratio of the slopes should recover the slope of the $M_\star - M_{\rm UV}$ relation. Therefore, I include a correction factor of 1.1 to equation 3.14 in order to make $\alpha_{\rm SM}$ and $\alpha_{\rm UV}$ comparable. The value 1.1 is based on the results in Table 3.1. If $M_\star \propto L$, this factor would become unity. I combine the measured biases over all redshifts, assume the same slope but different intercepts for each redshift, and fit these five parameters using the EMCEE MCMC sampler developed by Foreman-Mackey et al. (2013), assuming flat priors. Best-fit results are shown in the left and middle panels of Figure 3.4. I illustrate the marginalised distributions of the slopes for stellar mass and luminosity subsamples as solid red and blue lines respectively in the bottom left panel of Figure 3.4. The median and $1-\sigma$ percentiles of the slopes are summarised in the top right corner. I conclude that the slope in the stellar mass case is systematically smaller than in the luminosity case, which indicates larger clustering segregation with luminosity. The significance of

this trend can be seen from the bottom right panel of the figure. The probability that $\alpha_{\rm UV} > \alpha_{\rm SM}$ is $\simeq 90\%$ (shaded region).

However, UV magnitude only probes recent starbursts of a galaxy, while stellar mass is an integrated quantity, which reflects the whole star formation history. This argument suggests that a tighter relation is expected between stellar and halo mass, and therefore, clustering segregation with stellar mass should be larger. Both observational effects and astrophysical reasons could be responsible for this discrepancy. I investigate this issue further by comparing the results at $z \sim 4$ with predictions from the MERAXES semi-analytic model in the next section.

3.3 Comparison with Meraxes

3.3.1 Model overview

I use the MERAXES model introduced in Chapter 2 to compare the clustering measurements in the previous section. The model is run on the extended *Tiamat* N-body simulation (Poole et al., 2016, 2017). The simulation has 2160^3 dark matter particles, with particle mass $m_{\rm p} = 2.64 \times 10^6 h^{-1} M_{\odot}$, and side length $67.8\,h^{-1}$Mpc. The simulation outputs 101 snapshots between $z = 35$ and $z = 5$ with time steps separated by ~ 11 Myr, and additional 63 snapshots from $z = 5$ to $z = 1.8$. The time steps of these additional snapshots are evenly spaced in dynamical time. I adopt the same parameter configuration as in Qin et al. (2017) for MERAXES. All model stellar mass is subtracted by 0.24 dex in order to convert from a Salpeter (1955) IMF to a Chabrier (2003) IMF.

Photometry of model galaxies is calculated using the method described in Liu et al. (2016). I assume constant metallicity $Z = 0.001$ for each SSP. Recent hydrodynamical simulations (Ma et al., 2016; Davé et al., 2016) predict a mass-metallicity relation that is close to this value at $z \sim 4$. Cullen et al. (2017) point out that intrinsic luminosities only weakly depend on the metallicity. Therefore, this assumption is valid for this work. SSPs templates are generated by STARBURST99 Leitherer et al. (1999); Vázquez & Leitherer (2005); Leitherer et al. (2010, 2014) assuming a Salpeter (1955) IMF. Modelling the Lyman absorption of the intergalactic medium (IGM) is critical for the selection of model LBGs. I update the transmission curve of the IGM using a recent study from

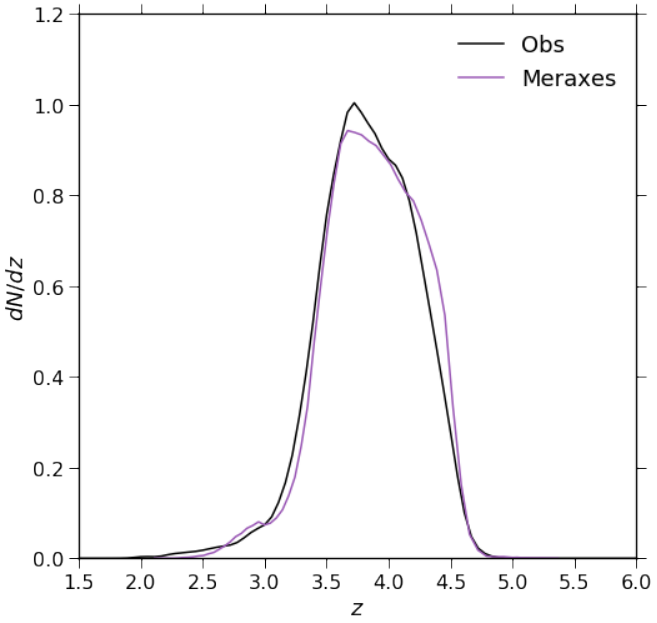


FIGURE 3.5: Example redshift distribution of observed and selected model galaxies in the $z \sim 4$ LBG sample. The observed distribution is estimated by photometric redshifts. The color selection on the model galaxies results in a very well-matched redshift distribution. The field depth of the XDF is used to select model galaxies.

Inoue et al. (2014), which predicts a more consistent redshift distribution for selected model LBGs. In addition, I also update the dust correction using an empirical model proposed by Mason et al. (2015), which is applicable at $z < 4$.

I follow Park et al. (2016, 2017) to select model LBGs and calculate the ACFs. This approach mimics the incompleteness of the LBG sample by adding photometric scatter to the magnitudes of each LBG selection band. The level of the scatter is given by the 1σ field detection limits. In the present work, I assemble model LBGs according to the flux limits of the deepest field in our observations, i.e. the XDF. The resulting redshift distribution of model selected LBGs is shown in Figure 3.5, which agrees with the observed one estimated by photometric redshifts. In terms of the determination of the ACF, I compute the real-space correlation function across a sequence of snapshots directly from the spatial coordinates of each galaxy, and convert to an ACF by the Limber transform. The readers are referred to Park et al. (2016, 2017) for a more detailed description.

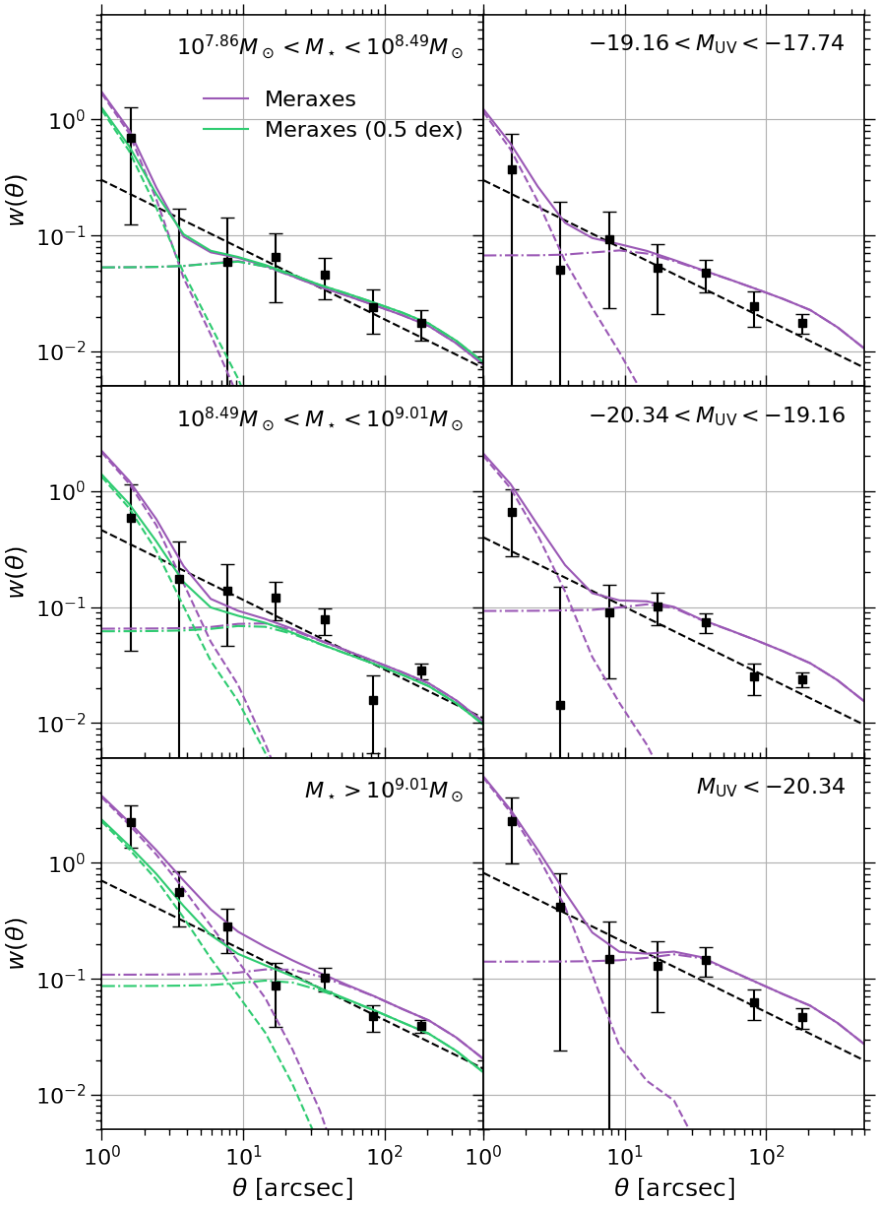


FIGURE 3.6: Comparison between observed and model predicted ACFs at $z \sim 4$. Black squares with error bars and black dash lines are observed ACFs and the corresponding best-fit power laws, which are the same with those in the top left panel of Figure 3.2. Purple lines are ACFs estimated from MERAXES. Purple dashed and dot dashed lines are the corresponding 1-halo and 2-halo terms, which is calculated by counting galaxy pairs in the same and different FoF groups. Green lines represent the model predictions from the sample where 0.5 dex Gaussian scatters are added to stellar mass.

3.3.2 Results

I focus on the comparison between our model and observations at $z \sim 4$, where the measurements have the smallest errors. I plot predicted ACFs together with measured ACFs in Figure 3.6. The model ACFs agree very well with observations, and reproduce the observed clustering dependence on both stellar mass and luminosity. In order to check whether the model predicts larger clustering segregation with luminosity as indicated by our observations, I therefore calculate the bias from the model by $b^2 = \xi(r)/\xi_{\rm DM}(r, z)$, where $\xi(r)$ and $\xi_{\rm DM}(r, z)$ are the real-space correlation functions of galaxies and dark matter. An average value is taken in the range $5\,{\rm Mpc} \leq r \leq 10\,{\rm Mpc}$. The estimated biases for the lightest (faintest) and heaviest (brightest) samples are shown using star symbols in the left (middle) panel in Figure 3.4. Subsequently, I derive the slope for these two points for comparison with observational data in Figure 3.4. I conclude that the measured variation of bias with stellar mass and luminosity is consistent with predictions and that the clustering segregation with luminosity is larger than stellar mass in the model. In other words, the model predictions also contradict my expectation that stellar mass should be more tightly correlated with halo mass.

The observed clustering dependence on stellar mass could be weakened due to observational uncertainties on the estimations of stellar mass. The uncertainties, for instance, can be due to the low S/N ratio of Spitzer data. I demonstrate this effect by adding 0.5 dex of Gaussian scatter to the stellar mass of selected model LBGs and remeasure the clustering in stellar mass bins. This level of scatter is also used in abundance matching studies (e.g. Moster et al., 2013; Behroozi et al., 2013). The recalculated ACFs are shown as green lines in Figure 3.6. For the most massive bin, the ACF decreases at all scales, while for the other two bins, scatter in stellar mass only affects the small scale correlation functions. I also recalculate the bias and the corresponding slope. The results are demonstrated in Figure 3.4. In the right panel of Figure 3.4, the dot dashed line represents the slope in the case where scatter is added to the stellar mass, and shows that scatter reduces the slope relative to that of the original model, from $\alpha_{\rm SM} = 1.3$ to $\alpha_{\rm SM} = 0.7$. This effect is most significant for the most massive bins, since the stellar mass function is steeper at the massive end. I can use this systematic error of model $\alpha_{\rm SM}$ introduced by adding scatter to the stellar mass to estimate the effect on the measured $\alpha_{\rm SM}$. This indicates that accounting for uncertainties in stellar mass leads to clustering

segregation that is similar for both mass and UV-luminosity, but which is not larger with stellar mass. Hence, although uncertainties in stellar mass can weaken the clustering dependence, the unexpected trend could still result from physical reasons.

Another interesting finding from the comparison between observations and the model is the deviation of the ACFs from a power law at small scales ($\theta < 10$ arcsec). In the model, this is due to the 1-halo term, arising from multiple halo occupation, where more than one galaxy resides in the same halo. To provide evidence of this multiple halo occupation, I calculate the 1-halo and 2-halo terms explicitly from MERAXES by counting galaxy pairs in the same and different FoF groups. These two terms are shown as dashed and dot dashed lines in Figure 3.6, respectively, demonstrating that the steep increase of the ACFs at small scales is due to multiple halo occupation. Consistency is also found between observations and the model. It can also be seen that the transition of the model ACFs between the 1-halo and 2-halo terms becomes more rapid with decreasing stellar mass. However, there is no such trend in luminosity. This finding may suggest an additional feature of clustering segregation with stellar mass and luminosity, implying different satellite properties for the two cases. However, I caution that the satellite properties of MERAXES have yet to be fully explored and compared with observations, and that achieving realistic recent satellite star formation histories has traitionally been a challenging task for semi-analytic models. This difference can also be seen from the observed ACFs but with large uncertainties. At the very bright end, the smooth transition between 1-halo and 2-halo terms is observed in Harikane et al. (2018), and explained using non-linear bias (Jose et al., 2016). Larger surveys with more complete samples might be used to investigate this phenomenon in more details for fainter and less massive galaxies.

3.4 Summary

I have carried out a clustering analysis of LBGs over the range $z \sim 4-7$, with emphasis on the comparison between clustering segregation with stellar mass and luminosity. I also compare the measurements with predictions from the MERAXES semi-analytic model. The main findings of this work are summarised as follows:

- The observed ACF amplitude and bias generally increase with stellar mass and luminosity over $z \sim 4-7$. The ACFs obtained from the model are consistent with observations, and reproduce clustering segregation with both stellar mass and luminosity. This suggests that more luminous and massive galaxies are more clustered, and hence hosted by more massive dark matter halos.

- By combining measurements over all redshifts, a systematic difference is found between clustering segregation with stellar mass and luminosity. In particular, it is observed that clustering strength is more tightly correlated with luminosity. This is in contrast to the expectation that stellar mass should be more tightly correlated with halo mass since stellar mass reflects the whole star formation history, while UV magnitude only corresponds to recent star formation. I find that the model also predicts this surprising result of larger clustering segregation with luminosity.

- At $z \sim 4$, the model predicts that the transition between the 1-halo and 2-halo terms of the ACFs is smoother for larger stellar mass galaxies, and that this trend does not appear for samples split by luminosity. Observations show similar behaviour but with large error bars. This might suggest that samples split by stellar mass and luminosity have quite different satellite properties.

Our results extend to higher redshift findings from the recent study at $z \sim 3$ by Durkalec et al. (2018), who carried out a clustering analysis of 3236 galaxies discovered in the VIMOS Ultra Deep Survey. They measured the real space correlation functions, and also found a larger difference in the correlation lengths when splitting the sample by luminosity than by stellar mass. This unexpected trend of clustering segregation with stellar mass and luminosity may provide new clues of galaxy formation in the early universe. For instance, some high-redshift galaxies might be formed by a single starburst. In this case, stellar mass and UV-luminosity become similar indicators of the star formation history. This work also motivates future spectroscopic surveys with the James Webb Space Telescope (JWST), which will provide more complete samples and more accurate stellar mass measurements substantially reducing the systematic errors in clustering studies.

Chapter 4

Dust extinction at high redshifts

4.1 Introduction

Having studied clustering segregation with UV-luminosity and stellar mass at $z \gtrsim 4$, I next investigate high redshift dust extinction in this chapter. In the early Universe, observations focus mainly on rest-frame UV properties due to cosmic redshift. These include measurements of UV luminosity functions (LFs) (van der Burg et al., 2010; Bouwens et al., 2015; Livermore et al., 2017; Bhatawdekar et al., 2018; Ono et al., 2018), and UV continuum slope to UV magnitude relations (Finkelstein et al., 2012; Bouwens et al., 2014; Rogers et al., 2014), which are also known as the colour-magnitude relations (CMRs). The UV luminosity is a tracer of star formation since most UV photons are emitted by young stars. However, star formation can be heavily obscured by the interstellar dust. One commonly adopted approach to perform dust corrections at high redshifts is to infer the infrared excess (IRX) from the observed UV slopes using a relation calibrated by Meurer et al. (1999) (e.g. Bouwens et al., 2015; Mason et al., 2015; Liu et al., 2016). However, the Meurer et al. (1999) relation is calibrated against local starburst galaxies, and observations of far infrared data are rather challenging at high redshifts. Recent observations at $z \gtrsim 3$ show large scatter in the IRX - β relation (Capak et al., 2015; Álvarez-Márquez et al., 2016; Bouwens et al., 2016; Barisic et al., 2017; Fudamoto et al., 2017; Koprowski et al., 2018). For instance, the observed IRX by Bouwens et al. (2016) is much lower than the Meurer et al. (1999) relation, while Koprowski et al. (2018) suggest that the IRX - β relation does not evolve with

redshift. These observations motivate investigation of the IRX - β at high redshifts from theoretical models.

Theoretical studies of dust extinction require intrinsic galaxy properties as input, and one approach is to postprocess the output of a hydrodynamical simulation. This method has been implemented in Safarzadeh et al. (2017) and Narayanan et al. (2018) to investigate the origin of the IRX - β relation. At $z \gtrsim 5$, the IRX - β relation has been studied by Mancini et al. (2016), Cullen et al. (2017) and Ma et al. (2019). However, their results suggest different extinction curves. Cullen et al. (2017) pointed out that the reason for the disagreement could be due to systematics associated with different simulations.

Predictions of semi-analytic models can also be used as input to dust models. Furthermore, semi-analytic models are computationally efficient, and therefore allow frequent exploration of different galaxy formation scenarios that produce different intrinsic galaxy luminosity. This work utilises the MERAXES semi-analytic model to predict intrinsic galaxy properties, and combines it with a simple and flexible dust attenuation model. The dust optical depths are calculated empirically using relevant galaxy properties. By taking full advantage of the fast computational speed of both the galaxy formation and dust models, I carry out a Bayesian analysis on all the model free parameters, and use UV LFs and CMRs as constraints, which are the most fundamental observables at high redshift. This approach allows these observations to put direct constraints on both galaxy formation and dust parameters, and provides self-consistent predictions of the IRX and star formation rate (SFR).

This chapter is organised as follows. Section 4.2 introduces several updates on MERAXES for this work. Section 4.3 describes the dust models that are integrated into MERAXES. I introduce the method to compute galaxy spectral energy distributions (SEDs) in Section 4.4. The description of my calibration method can found in Section 4.5, and the results are discussed in Section 4.6. I demonstrate the predicted IRX - β relations and cosmic star formation rate density (SFRD) in Section 4.7 and Section 4.8 respectively. Finally, this work is summarised in Section 4.9. Throughout the paper, weadopt a flat ΛCDM cosmology, with $(h, \Omega_{\rm m}, \Omega_{\rm b}, \Omega_\Lambda, \sigma_8, n_{\rm s}) = (0.678, 0.308, 0.0484, 0.692, 0.815, 0.968)$ (Planck Collaboration et al., 2016). Magnitudes are in the AB system (Oke & Gunn, 1983).

4.2 Updates to Meraxes

This works improves the supernova feedback model of MERAXES in Mutch et al. (2016) by adopting different mass loading factor η and supernova energy coupling efficiency:

$$\eta = \begin{cases} \eta_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm reheat}} \left(\frac{V_{\rm max}}{60{\rm km/s}}\right)^{-1}, & V_{\rm max} \geq 60{\rm km/s} \\ \eta_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm reheat}} \left(\frac{V_{\rm max}}{60{\rm km/s}}\right)^{-3.2}, & V_{\rm max} < 60{\rm km/s} \end{cases}, \tag{4.1}$$

$$\epsilon = \begin{cases} \epsilon_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm eject}} \left(\frac{V_{\rm max}}{60{\rm km/s}}\right)^{-1}, & V_{\rm max} \geq 60{\rm km/s} \\ \epsilon_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm eject}} \left(\frac{V_{\rm max}}{60{\rm km/s}}\right)^{-3.2}, & V_{\rm max} < 60{\rm km/s} \end{cases}, \tag{4.2}$$

where $V_{\rm max}$ is the maximum circular velocity. Muratov et al. (2015) originally obtained a broken power law for the mass loading factor. Their study is based on model galaxies in the FIRE simulations (Hopkins et al., 2014). This form is subsequently implemented in several semi-analytic models (Hirschmann et al., 2016; Cora et al., 2018; Lagos et al., 2018). The implementation of this form in the present work is primarily motivated by its impact on the metallicity, which is an input of galaxy SEDs. Hirschmann et al. (2016) tested eight different supernova feedback schemes in their semi-analytic model, and found that only explicit redshift-dependent models can lead to evolution of the mass metallicity relation. Collacchioni et al. (2018) demonstrated that a steeper slope of the redshift dependence can result in stronger evolution of the mass metallicity relation using the semi-analytic model of Cora et al. (2018). In this work, I set $\alpha_{\rm reheat} = 2$ according to the optimisation result in Cora et al. (2018), assume no redshift dependence on the energy coupling efficiency (i.e. $\alpha_{\rm eject} = 0$) and leave η_0 and ϵ_0 as free parameters.

In addition, this work evaluates $d\epsilon/d\tau$ and $dy/d\tau$ in Equations 2.12, 2.13, and 2.16 using STARBURST99 (Leitherer et al., 1999; Vázquez & Leitherer, 2005; Leitherer et al., 2010, 2014) with a Kroupa (2002) initial mass function (IMF). This treatment provides more reasonable and self-consistent estimates of these quantities, and can be generalised to other stellar evolutionary libraries (e.g Saitoh, 2017; Ritter et al., 2018). A similar approach has already been applied in the FIRE hydrodynamic simulations (Hopkins et al., 2014).

4.3 Dust models

I implement the dust model proposed by (Charlot & Fall, 2000). The transmission function due to the ISM is expressed by

$$T_\lambda(t) = \begin{cases} \exp(-\tau_\lambda^{\rm ISM}) & t \geq t_{\rm BC} \\ \exp(-\tau_\lambda^{\rm ISM} - \tau_\lambda^{\rm BC}) & t < t_{\rm BC} \end{cases}. \tag{4.3}$$

This model takes into account the relative stars-dust geometry of different stellar populations. Photons emitted by young stars are absorbed by an additional component due to the surrounding molecular cloud where the stars form. The birthcloud is assumed to have lifetime $t_{\rm BC}$, and for stars whose age is older than $t_{\rm BC}$, their starlight is only absorbed by the diffuse ISM dust. I fix $t_{\rm BC} = 10$ Myr according to previous studies (Charlot & Fall, 2000; da Cunha et al., 2008). The attenuation due to the birth cloud and diffuse ISM dust is described by their optical depths $\tau_\lambda^{\rm BC}$ and $\tau_\lambda^{\rm ISM}$ respectively, which should vary with different galaxies. In this study, I explore three different parametrisations, linked to star formation rate (SFR), dust-to-gas (DTG) ratio and gas column density (GCD). I name them as M-SFR, M-DTG and M-GCD respectively. In general, these properties are indirectly related to the dust. One dust production channel is from the ejecta of supernova (e.g. Dayal & Ferrara, 2018), which is proportional to the SFR. Dust is also mixed with gas. Accordingly, they are expected to have similar properties. I will see that M-DTG and M-GCD have similar results since they primarily depend on gas density.

4.3.1 Star formation rate model

The dependence of the dust optical depths on SFR is motivated by observations of the CMRs at high redshifts, i.e the relation between UV continuum slope and UV magnitude. These observations suggest that more UV luminous galaxies have redder UV continuum slopes (Finkelstein et al., 2012; Bouwens et al., 2014; Rogers et al., 2014). Since brighter galaxies correspond to higher SFR, one could expect that SFR and dust content are positively correlated. Similar trends have been found in low redshift studies (e.g. da

Cunha et al., 2010; Qin et al., 2019a). Hence, I assume the following parameterisation

$$\Gamma_\lambda = e^{-az} \left(\frac{\mathrm{SFR}}{100\,\mathrm{M}_\odot/\mathrm{yr}} \right)^{\gamma_{\mathrm{SFR}}} \left(\frac{\lambda}{1600\,\text{Å}} \right)^n, \tag{4.4}$$

$$\tau_\lambda^{\mathrm{ISM}} = \tau_{\mathrm{SFR}}^{\mathrm{ISM}} \Gamma_\lambda, \tag{4.5}$$

$$\tau_\lambda^{\mathrm{BC}} = \tau_{\mathrm{SFR}}^{\mathrm{BC}} \Gamma_\lambda, \tag{4.6}$$

where $\tau_{\mathrm{SFR}}^{\mathrm{ISM}}$, $\tau_{\mathrm{SFR}}^{\mathrm{ISM}}$, γ_{SFR}, a and n are free parameters. Yung et al. (2019) also use a parametric model to calculate dust attenuation in their semi-analytic model. They adjusted the normalisation of the optical depth to fit the observed UV LFs at individual redshifts. Their results indicate that the normalisation depends on redshift and the trend can be fit by an exponential function. Therefore, for all the three parametrisations proposed in this work, I also include an exponential redshift dependence factor to fit the model against multiple redshifts.

4.3.2 Dust-to-gas ratio model

In the literature, dust optical depths are often linked to the gas column density, which is then converted to the dust column density using the dust-to-gas (DTG) ratio (De Lucia & Blaizot, 2007; Guo et al., 2011; Somerville et al., 2012; Yung et al., 2019). In this model, optical depths are expressed by

$$\Gamma_\lambda = e^{-az} \left(\frac{Z_{\mathrm{cold}}}{Z_\odot} \right)^{\gamma_{\mathrm{DTG}}} \left(\frac{m_{\mathrm{cold}}}{10^{10} h^{-1}\,\mathrm{M}_\odot} \right) \left(\frac{r_{\mathrm{disk}}}{h^{-1}\,\mathrm{kpc}} \right)^{-2} \left(\frac{\lambda}{1600\,\text{Å}} \right)^n, \tag{4.7}$$

$$\tau_\lambda^{\mathrm{ISM}} = \tau_{\mathrm{DTG}}^{\mathrm{ISM}} \Gamma_\lambda, \tag{4.8}$$

$$\tau_\lambda^{\mathrm{BC}} = \tau_{\mathrm{DTG}}^{\mathrm{BC}} \Gamma_\lambda, \tag{4.9}$$

where Z_{cold} is the metallicity of cold gas, m_{cold} is the mass of cold gas, and r_{disk} is the disk scale radius defined in Equation (2.7). I adopt the solar metallicity as $Z_\odot = 0.02$. Free parameters are $\tau_{\mathrm{DTG}}^{\mathrm{ISM}}$, $\tau_{\mathrm{DTG}}^{\mathrm{BC}}$, γ_{DTG}, a and n.

4.3.3 Gas column density model

I propose an additional gas mass related dust model, which is independent of the metallicity. In Mutch et al. (2016) and this work, when metals are produced by supernova

explosions, I assume that they are first fully mixed with cold gas, and then ejected into the hot gas reservoir. In reality, since the materials produced by supernova have quite different initial velocities from the surrounding gas, the mixing may take some time. Thus, I provide a metallicity independent parametrisation of the dust optical depths

$$\Gamma_\lambda = e^{-az} \left(\frac{m_{\rm cold}}{10^{10} h^{-1}\ {\rm M}_\odot} \right)^{\gamma_{\rm GCD}} \left(\frac{r_{\rm disk}}{h^{-1}\,{\rm kpc}} \right)^{-2} \left(\frac{\lambda}{1600\,{\rm \AA}} \right)^{n}, \tag{4.10}$$

$$\tau_\lambda^{\rm ISM} = \tau_{\rm GCD}^{\rm ISM} \Gamma_\lambda, \tag{4.11}$$

$$\tau_\lambda^{\rm BC} = \tau_{\rm GCD}^{\rm BC} \Gamma_\lambda. \tag{4.12}$$

There are also five free parameters in this model, i.e. $\tau_{\rm GCD}^{\rm ISM}$, $\tau_{\rm GCD}^{\rm BC}$, $\gamma_{\rm GCD}$, a and n. This model includes a power law scaling on the cold gas mass, unlike the M-DTG model, where the scaling is on metallicity.

4.4 Synthetic spectral energy distributions

The computation of galaxy spectral energy distributions (SEDs) follows standard stellar population synthesis. The luminosity of a galaxy at time t can be obtained by

$$L_\lambda(t) = \int_0^t d\tau \int_{Z_{\rm min}}^{Z_{\rm max}} dZ\, \psi(t-\tau, Z) S_\lambda(\tau, Z) T_\lambda(\tau), \tag{4.13}$$

where τ is the stellar age, $\psi(t-\tau, Z)\, d\tau dZ$ is the mass of stars formed at $t-\tau$ with an age between τ to $\tau + d\tau$ and metallicity between Z to $Z + dZ$, $S_\lambda(\tau, Z)$ is the luminosity of a simple stellar population (SSP) per unit mass, and $T_\lambda(\tau)$ is the transmission function of the ISM described in the previous subsection. I generate $S_\lambda(\tau, Z)$ using STARBURST99 (Leitherer et al., 1999; Vázquez & Leitherer, 2005; Leitherer et al., 2010, 2014), assuming a metallicity range from $Z = 0.001$ to $Z = 0.040$ and a Kroupa (2002) IMF. Nebular continuum emssions are also added using STARBURST99. To compute UV magnitudes, I apply a tophat filter centred at $\lambda = 1600$Å with width 100Å. UV slopes are obtained by a linear fit in the logarithmic flux space using the ten windows proposed by Calzetti et al. (1994). However, for computational speed, I only choose five of them (including the longest wavelength window) for on-the-fly calibrations. The selected windows are given in Table 4.1. The median errors from this treatment are negligible in the range of the observed CMRs.

	Wavelength range [Å]
1	1342 - 1371
2	1562 - 1583
3	1866 - 1890
4	1930 - 1950
5	2400 - 2580

TABLE 4.1: The five windows selected from Calzetti et al. (1994) to fit UV slopes for the on-the-fly calibrations.

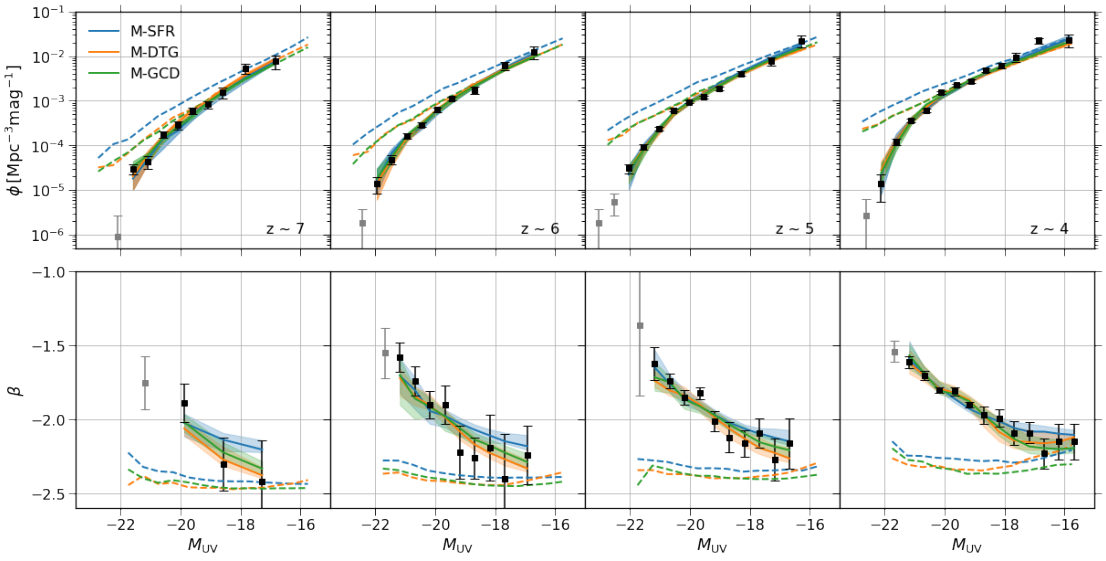


FIGURE 4.1: Best-fit luminosity functions (LFs) and colour-magnitude relations (CMRs). Solid blue, orange and green lines are the results of M-SFR (Section 4.3.1), M-DTG (Section 4.3.2), M-GCD (Section 4.3.3) respectively. Shaded regions illustrate the 1σ (68 %) range of the posterior distributions. Dashed lines are the corresponding dust-unattenuated properties. Black points with errorbars are the observational data used in the calibration, which are from Bouwens et al. (2015) and Bouwens et al. (2014) for the LFs and CMRs respectively. Grey data points are also from these observations but are not used in the calibration due to the limit of the simulation box size.

I also make a numeric approximation in order to accelerate the speed of evaluating Equation (4.13). I first compute the intrinsic luminosity in necessary filters. The dust transmission is then applied to the luminosity of the filters using the central wavelength instead of the full SEDs. This approximation is found to have a negligible effect on the results, since all filters used in this work have a simple shape and are relatively narrow.

TABLE 4.2: Summary of free galaxy and dust parameters.

Parameter	Section	Equation	Description
$\alpha_{\rm SF}$	Sec.2.3	Eq.2.8	Star formation efficiency
$\Sigma_{\rm SF}$	Sec.2.3	Eq.2.6	Critical mass normalisation
η_0	Sec.4.2	Eq.4.1	Mass loading normalisation
ϵ_0	Sec.4.2	Eq.4.2	supernova energy coupling normalisation
$\tau^{\rm ISM}_{\rm SFR}/\tau^{\rm ISM}_{\rm DTG}/\tau^{\rm ISM}_{\rm GCD}$	Sec.4.3.1/Sec.4.3.2/Sec.4.3.3	Eq.4.5/Eq.4.8/Eq.4.11	Dust optical depth normalisation of ISM
$\tau^{\rm BC}_{\rm SFR}/\tau^{\rm BC}_{\rm DTG}/\tau^{\rm BC}_{\rm GCD}$	Sec.4.3.1/Sec.4.3.2/Sec.4.3.3	Eq.4.6/Eq.4.9/Eq.4.12	Dust optical depth normalisation of BC
$\gamma_{\rm SFR}/\gamma_{\rm DTG}/\gamma_{\rm GCD}$	Sec.4.3.1/Sec.4.3.2/Sec.4.3.3	Eq.4.4/Eq.4.7/Eq.4.10	Dust optical depth slope of galaxy property
n	Sec.4.3.1/Sec.4.3.2/Sec.4.3.3	Eq.4.4/Eq.4.7/Eq.4.10	Reddening slope
a	Sec.4.3.1/Sec.4.3.2/Sec.4.3.3	Eq.4.4/Eq.4.7/Eq.4.10	Dust optical depth redshift dependence

Parameter	Prior scale	Prior range			Best-fit [a]			16/84-th percentiles [b]		
		M-SFR	M-DTG	M-GCD	M-SFR	M-DTG	M-GCD	M-SFR	M-DTG	M-GCD
$\alpha_{\rm SF}$	log	[0.005, 0.2]	[0.05, 0.18]	[0.04, 0.08]	0.10	0.10	0.05	[0.08, 0.13]	[0.10, 0.11]	[0.05, 0.07]
$\Sigma_{\rm SF}$	log	[0.1, 0.8]	[0.001, 0.25]	[0.05, 0.25]	0.19	0.01	0.16	[0.21, 0.42]	[0.007, 0.06]	[0.14, 0.19]
η_0	log	[2.0, 12.0]	[2.0, 15.0]	[3.5, 7.5]	4.6	7.0	6.4	[4.0, 7.8]	[6.6, 7.9]	[4.9, 6.1]
ϵ_0	log	[0.35, 0.65]	[0.8, 2.2]	[1.0, 1.7]	0.5	1.5	1.3	[0.4, 0.6]	[1.5, 1.7]	[1.3, 1.5]
$\tau^{\rm ISM}_{\rm SFR}/\tau^{\rm ISM}_{\rm DTG}/\tau^{\rm ISM}_{\rm GCD}$	linear	[0.5, 2.4]	[0.0, 50.0]	[2.0, 8.0]	1.7	13.5	3.7	[1.4, 1.7]	[9.9, 17.0]	[3.5, 5.3]
$\tau^{\rm BC}_{\rm SFR}/\tau^{\rm BC}_{\rm DTG}/\tau^{\rm BC}_{\rm GCD}$	linear	[2.0, 10.0]	[0.0, 1000.0]	[25.0, 140.0]	2.5	381.3	69.7	[3.9, 6.6]	[225.1, 476.1]	[60.4, 91.0]
$\gamma_{\rm SFR}/\gamma_{\rm DTG}/\gamma_{\rm GCD}$	linear	[0.0, 0.6]	[0.4, 2.2]	[1.3, 1.7]	0.19	1.20	1.48	[0.23, 0.32]	[1.05, 1.38]	[1.44, 1.52]
n	linear	[-1.00, -0.25]	[-2.5, -0.8]	[-1.6, -1.0]	-0.3	-1.6	-1.3	[-0.5, -0.3]	[-1.7, -1.5]	[-1.4, -1.2]
a	linear	[0.00, 0.15]	[0.10, 0.65]	[0.20, 0.55]	0.04	0.34	0.39	[0.02, 0.07]	[0.25, 0.37]	[0.36, 0.42]

[a]Sample point that has the highest posterior distribution value are chosen to be the best-fit values.
[b]These are the 16-th and 84-th percentiles of the marginalised distributions.

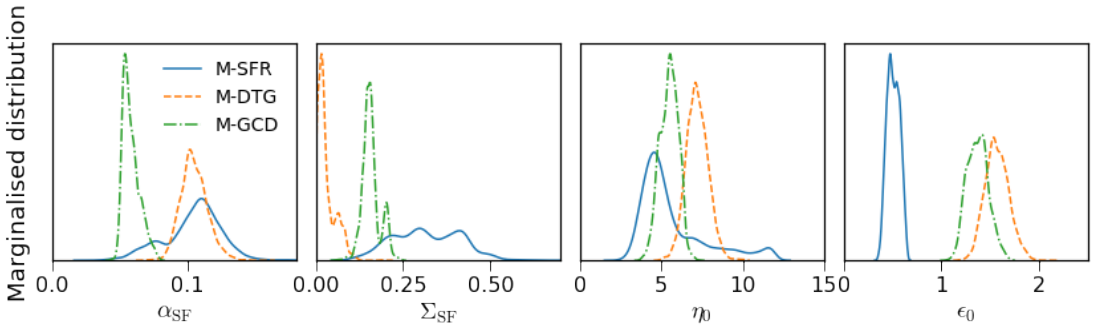


FIGURE 4.2: Comparison of the marginalised distributions of galaxy formation parameters among the three different dust models. The parameters are the star formation efficiency $\alpha_{\rm SF}$ (Equation 2.8), the normalisation of the critical mass $\Sigma_{\rm SF}$ (Equation 2.6), the mass loading factor η_0 (Equation 4.1) and the supernova energy coupling efficiency ϵ_0 (Equation 4.2). The three dust models labelled as M-SFR, M-DTG and M-GCD are described in Section 4.3.1, 4.3.2 and 4.3.3, and the corresponding optical depths in the three models are linked to the star formation rate (SFR), dust to gas ratio (DTG) and gas column density (GCD) respectively. The y-axes show the probability distributions in a linear scale.

4.5 Calibration

An essential part of this work is to determine the free parameters in both the galaxy formation and dust attenuation models introduced in the previous sections. I carry out a Bayesian analysis on these parameters, and use observed UV LFs and CMRs at $z \sim 4-7$ as constraints.

A key goal of a Bayesian analysis is to estimate the posterior distribution of model parameters, which is non-trivial for high dimensional spaces. Kampakoglou et al. (2008) and Henriques et al. (2009) first applied the Markov chain Monte Carlo (MCMC) method to sample the parameter space of semi-analytic models. This approach has been implemented by several subsequent studies (Henriques et al., 2013; Mutch et al., 2013; Henriques et al., 2015). However, the MCMC method has several drawbacks. Firstly, it requires additional evaluations of the model to ensure the final sample reaches a stationary distribution, and it is generally difficult to determine whether a Monte Carlo chain has fully converged (see Cowles & Carlin, 1996, for a review). Moreover, without special treatments, MCMC samplers can encounter difficulties in approaching a stationary distribution when the parameter space contains isolated modes (which is the case in this study), since random walkers can be trapped by a local minimum and fail to jump to other modes (e.g Neal, 1996). A possible improvement to handle multimodal distributions for MCMC methods can be found in Earl & Deem (2005).

In this work, in order to achieve higher sampling efficiency and obtain more stable results on multimodal parameter spaces, I utilise the multimodal nested sampling introduced by Feroz et al. (2009) to estimate the posterior distributions. This algorithm is found to be a competitive alternative to MCMC methods, and addresses the issues mentioned above to some extent. The nested sampling was designed to evaluate the Bayesian evidence (Skilling, 2004). However, the output samples produced by the algorithm can also be used to estimate posterior distributions, which is equivalent to the MCMC method. In difference from MCMC methods, no burn-in phase is required in this algorithm. The stopping criterion of the nested sampling is based on an estimated error of the resulting value of the Bayesian evidence, which is also proposed by Skilling (2004). The sampling efficiency of the original algorithm is improved by Feroz et al. (2009), who use the information of existing sample points to approximate the iso-likelihood surfaces in the parameter space as hyper ellipsoids (see also Mukherjee et al., 2006). Secondly, the algorithm includes a special treatment for multimodal problems. Again using the information of existing sample points, it applies a clustering algorithm to detect multiple modes and splits the parameter space (see also Shaw et al., 2007). This approach has been tested against toy models which contain several equally-high peaks, and is found to have good performance. The reader is referred to Feroz & Hobson (2008), Feroz et al. (2009) and references therein for a detailed description of the algorithm. A comparison between the nested sampling and the MCMC method can be found in Speagle (2019).

The Bayesian posterior distribution is comprised of the likelihood and prior distributions of each free model parameter. I construct the log-likelihood as

$$\begin{aligned}\ln \mathcal{L} = &-\frac{1}{2}\sum_i \left[\frac{(n_i^{\rm obs}-n_i^{\rm model})^2}{\sigma^2_{\rm LF,\,i}} + \ln(2\pi\sigma^2_{\rm LF,\,i})\right] \\ &-\frac{1}{2}\sum_i \left[\frac{(\beta_i^{\rm obs}-\beta_i^{\rm model})^2}{\sigma^2_{\rm CMR,\,i}} + \ln(2\pi\sigma^2_{\rm CMR,\,i})\right].\end{aligned} \tag{4.14}$$

Observational data of LFs ($n_i^{\rm obs}$, $\sigma^2_{\rm LF,\,i}$) and CMRs ($\beta_i^{\rm obs}$, $\sigma^2_{\rm CMR,\,i}$) are taken from Bouwens et al. (2015) and the biweight mean measurements of Bouwens et al. (2014) respectively. The LFs are defined by the co-moving number density. I convert the dimensionless Hubble constant from $h = 0.7$ to $h = 0.678$ for these observations in order to be consistent with my model. Due to the limited size of the simulation box, the model is unable to probe the full range of the LFs and CMRs. Therefore, for each LF

and CMR bin, I use the observed LF to estimate an expected number of galaxies in the simulation box, and drop the bin if the number is less than five and twenty for the LF and CMR respectively.

The model parameters are from both the semi-analytic model and the dust relations introduced in Section 4.3. I focus on four galaxy formation parameters: the star formation efficiency $\alpha_{\rm SF}$, normalisation of the critical mass $\Sigma_{\rm SF}$, mass loading factor η_0, and supernova energy coupling efficiency ϵ_0. Their prior distributions are chosen to be uniform in logarithmic space. Three different dust models were introduced in Section 4.3. Each of them has five free parameters. I adopt uniform priors in linear space for them.

The prior ranges of these model parameters are used in the initialisation of nested sampling, and they are chosen based on several experiments. I first run the sampler in a very large parameter space and find the high probability regions. I then shrink the prior ranges accordingly, keeping the posterior distribution at the bounds negligible compared with the high probability regions. There is an exception for the mass loading factor η_0 in the M-SFR model, which is found to have no upper limit. However, this will not affect my main results, since the energy coupling efficiency ϵ_0 puts a physical upper limit on the strength of the supernova feedback, and this parameter is constrained. This approach of choosing the prior ranges allows the sampler to spend more time on the high probability regions and improves the sampling efficiency. A summary of all model parameters and their prior ranges can be found in Table 4.2.

In practice, I utilise a modified version of the open source Python package NESTLE[1], which implements the algorithm, and couple it with the MERAXES Python interface MHYSA (Mutch, in prep.). I set the number of active points to be 300 for the sampler. The stop criterion follows the remaining Bayesian evidence approach suggested by Skilling (2004). The algorithm terminates when the logarithmic change due to the remaining Bayesian evidence is below one, and the convergence requires evaluating the model for 50,000 - 100,000 times.

[1]https://github.com/kbarbary/nestle. See https://github.com/smutch/nestle for the modified version.

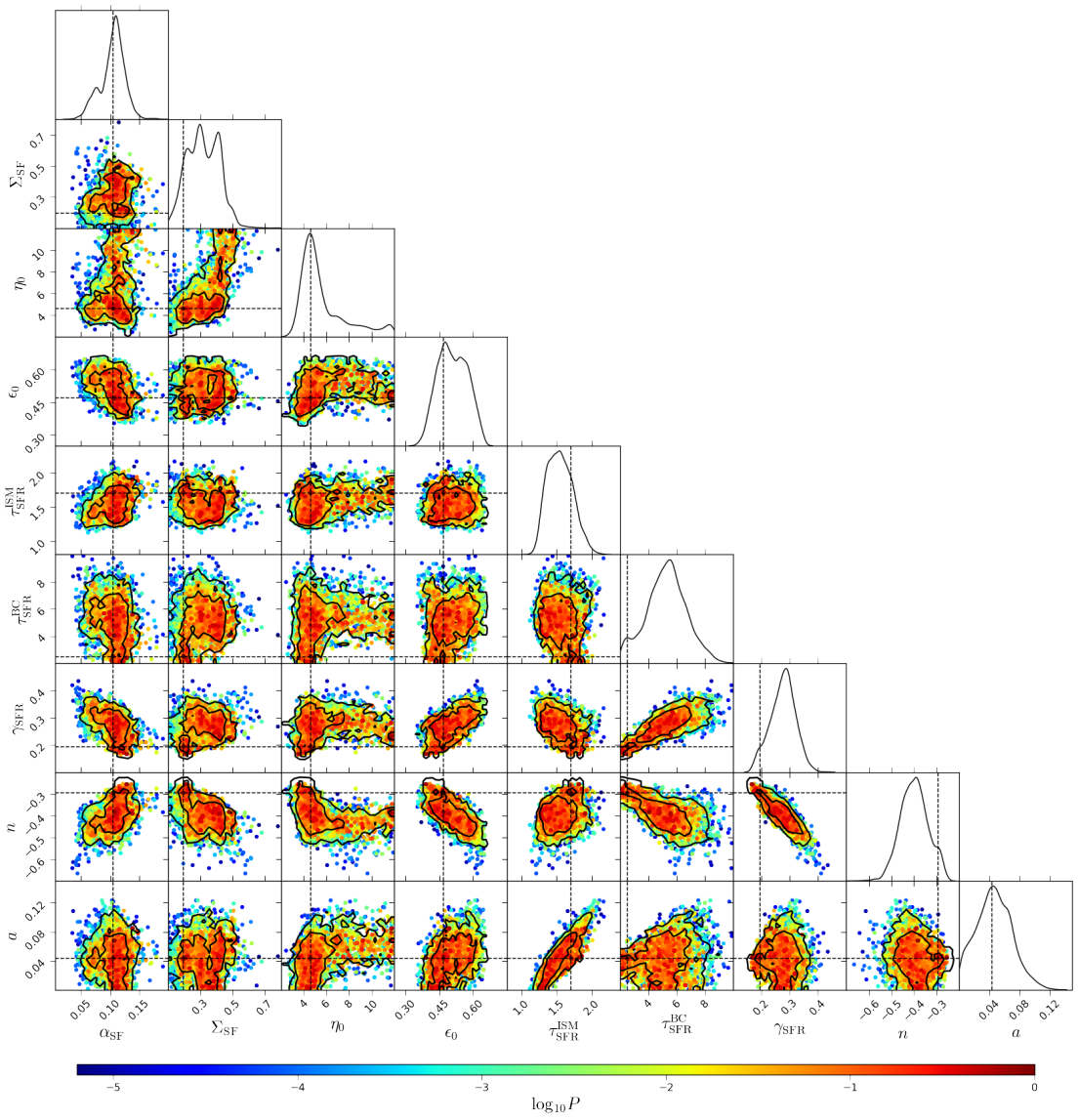


FIGURE 4.3: Posterior distribution of the galaxy and dust parameters for MERAXES with a star formation rate dependent (SFR) dust model. The model is referred to as M-SFR, which is described in Section 4.3.1. The posterior distribution is a function of star formation efficiency $\alpha_{\rm SF}$, critical mass normalisation $\Sigma_{\rm SF}$, mass loading factor η_0, supernova energy coupling efficiency ϵ_0, optical depth normalisations of interstellar media $\tau_{\rm SFR}^{\rm ISM}$ and birth cloud $\tau_{\rm SFR}^{\rm BC}$, optical depth scaling of star formation rate $\gamma_{\rm SFR}$, reddening slope n and optical depth redshift dependence a. See also Table 4.2 for a summary of these parameters. Diagonal panels show the one parameter marginalised distributions. In the off-diagonal panels, solid black lines are the 68% and 95% contours of the two parameter marginalised distributions. Colour points reflect the values of the posterior distribution, and the maximum is normalised to unity. The point that has the highest value is chosen to be best-fit results, which is specified by the dashed lines. Their values are listed in Table 4.2.

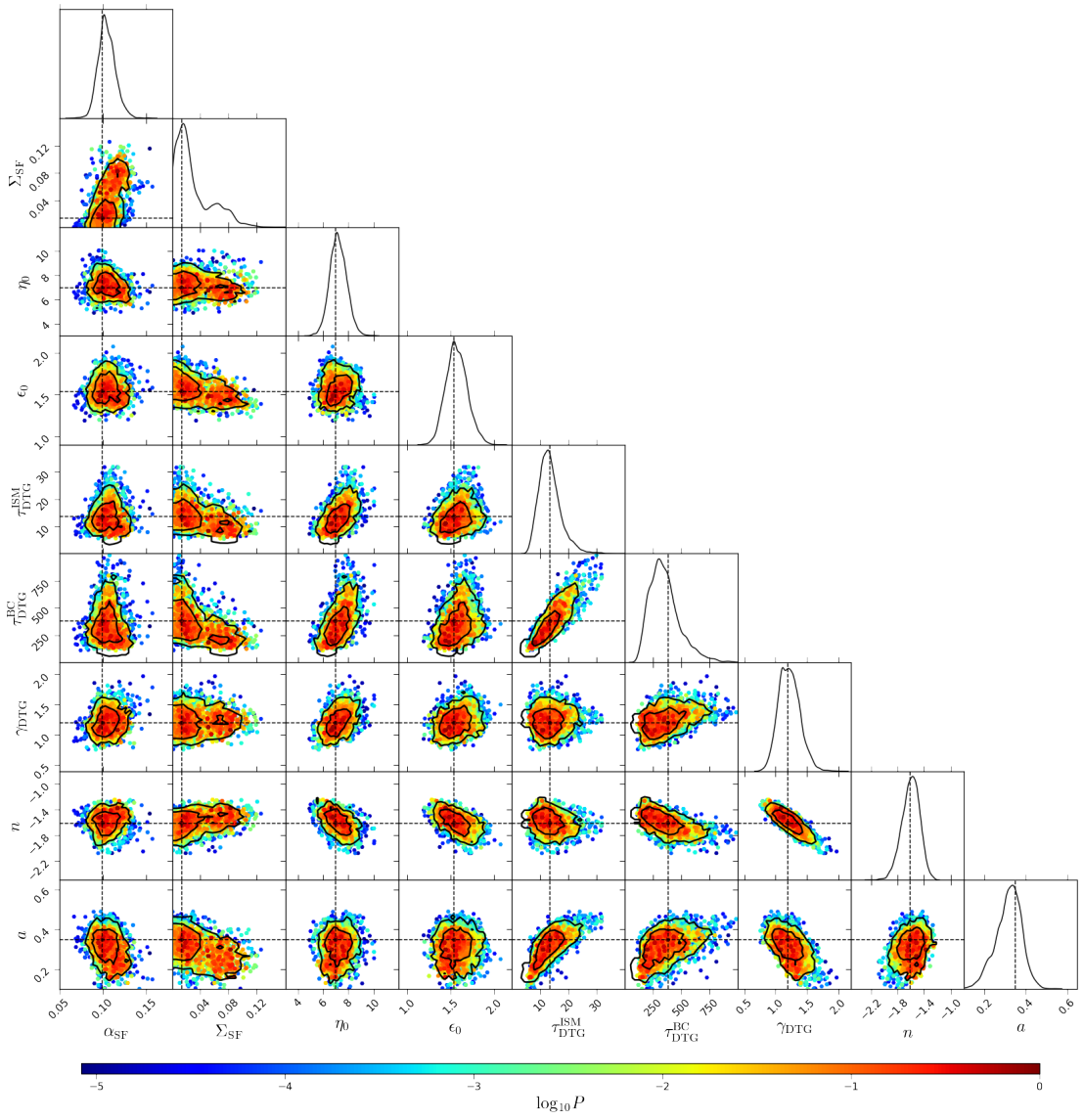


FIGURE 4.4: Posterior distribution of the galaxy and dust parameters for MERAXES with a dust-to-gas ratio (DTG) dependent dust model. The model is referred to as M-DTG and described in Section 4.3.2. The posterior distribution is a function of star formation efficiency $\alpha_{\rm SF}$, critical mass normalisation $\Sigma_{\rm SF}$, mass loading factor η_0, supernova energy coupling efficiency ϵ_0, optical depth normalisations of interstellar media $\tau_{\rm DTG}^{\rm ISM}$ and birth cloud $\tau_{\rm DTG}^{\rm BC}$, slope of the dust-to-gas ratio $\gamma_{\rm DTG}$, reddening slope n and optical depth redshift dependence a. See also Table 4.2 for a summary of these parameters. Diagonal panels show the one parameter marginalised distributions. In the off-diagonal panels, solid black lines are the 68% and 95% contours of the two parameter marginalised distributions. Colour points reflect the values of the posterior distribution, and the maximum is normalised to unity. The point that has the highest value is chosen to be best-fit results, which is specified by the dashed lines. Their values are listed in Table 4.2.

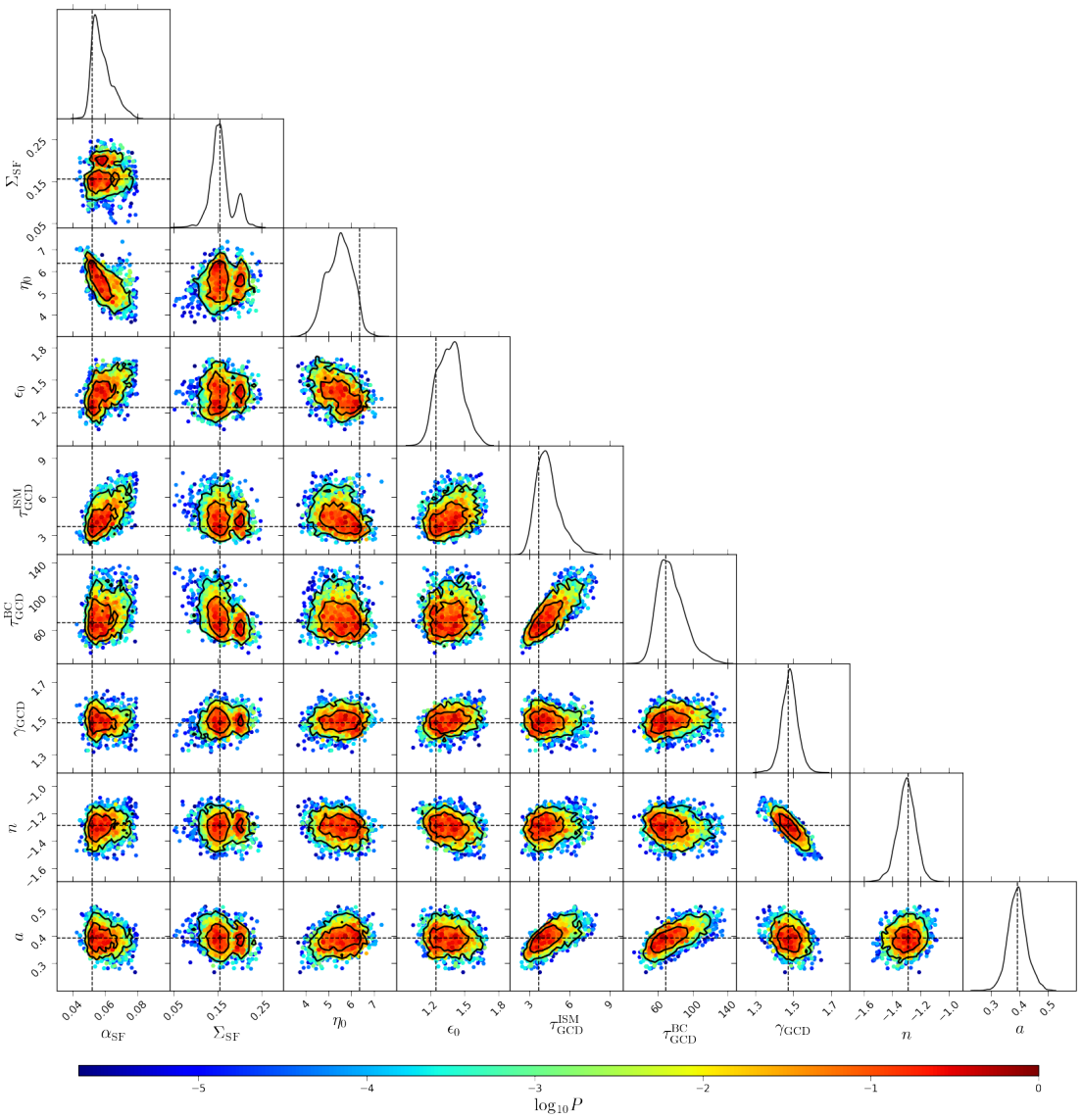


FIGURE 4.5: Posterior distribution of the galaxy and dust parameters for MERAXES with a gas column density (GCD) dependent dust model. The model is referred to as M-GCD and described in Section 4.3.3. The posterior distribution is a function of star formation efficiency $\alpha_{\rm SF}$, critical mass normalisation $\Sigma_{\rm SF}$, mass loading factor η_0, supernova energy coupling efficiency ϵ_0, optical depth normalisations of interstellar media $\tau_{\rm GCD}^{\rm ISM}$ and birth cloud $\tau_{\rm GCD}^{\rm BC}$, optical depth scaling of gas mass $\gamma_{\rm GCD}$, reddening slope n and optical depth redshift dependence a. See also Table 4.2 for a summary of these parameters. Diagonal panels show the one parameter marginalised distributions. In the off-diagonal panels, solid black lines are the 68% and 95% contours of the two parameter marginalised distributions. Colour points reflect the values of the posterior distribution, and the maximum is normalised to unity. The point that has the highest value is chosen to be best-fit results, which is specified by the dashed lines. Their values are listed in Table 4.2.

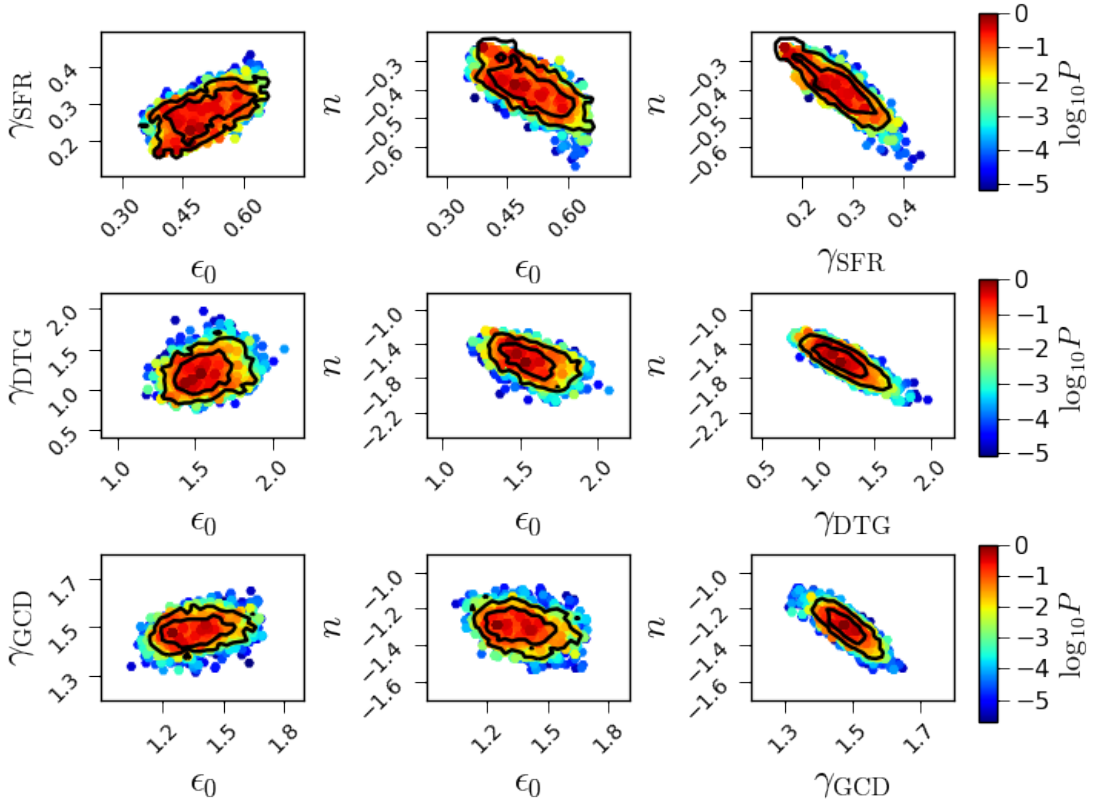


FIGURE 4.6: Correlations among the supernova energy coupling efficiency ϵ_0, galaxy property scaling of the dust relation $\gamma_{\rm SFR,\,DTG,\,GCD}$ and the reddening slope n. In each panel, solid back lines are the 68% and 95% contours of the two parameter marginalised distributions. Colour points indicate the values of the corresponding posterior distributions, and their maximum is normalised to unity. From top to bottom, rows correspond to the dust attenuation model of M-SFR (Section 4.3.1), M-DTG (Section 4.3.2), M-GCD (Section 4.3.3) respectively. The posterior distributions of all parameters for the three models can be seen in Figures 4.3, 4.4 and 4.5.

4.6 Fitting results

For the three different dust models, I obtain 5,000 - 6,000 sample points from the nested sampling algorithm, which describe the posterior distributions of both galaxy and dust parameters. The point that has the highest value of the posterior distribution is chosen to be the best-fit result. The best-fit parameter values are listed in Table 4.2, and the corresponding LFs and CMRs are shown in Figure 4.1 for each dust model. The three models all fit the observational data extremely well. In Figures 4.3, 4.4 and 4.5, I show the posterior distributions of M-SFR, M-DTG and M-GCD respectively. In plotting these figures, I adopt a similar approach with Henriques et al. (2009) and Henriques et al. (2013), i.e. using contours to show the marginalised distributions and colours to indicate the values of the whole posterior distributions. For the M-SFR model, it can

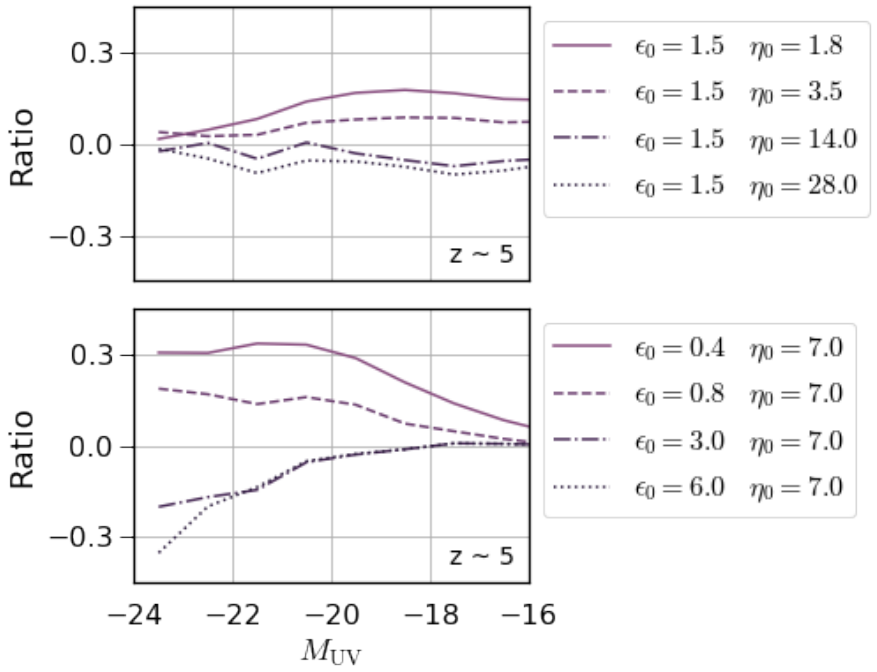


FIGURE 4.7: Effects of varying the mass loading factor η_0 and the supernova energy coupling efficiency ϵ_0 on the intrinsic UV luminosity function. The y-axes show the ratio of the logarithmic luminosity functions between the model variants and the best-fit M-DTG models. In the upper panel, I vary the mass loading factor η_0 at fixed ϵ_0, while in the lower panel, I fix ϵ_0 and change η_0.

be seen from Figure 4.2 that the marginalised distribution of the mass loading factor η_0 extends to large values, which means that this parameter is less constrained. On the other hand, all parameters are well constrained for the other two models.

An interesting finding is that the derived galaxy formation parameters preferred by these three dust models are quite different. Figure 4.2 illustrates a comparison of the marginalised distributions for the four galaxy formation parameters. I found that M-DTG and M-GCD suggest similar mass loading factor and supernova energy coupling efficiency. However, M-DTG shows evidence of a more active star formation scenario, with higher star formation efficiency and lower normalisation of the critical mass. For M-SFR, the marginalised distribution of the star formation efficiency overlaps with that of M-DTG. However, M-SFR requires much smaller supernova energy coupling efficiency. Moreover, differences can also be found in the two parameter correlations between the galaxy formation parameters for the three different models. For instance, in the third row and first column of Figure 4.5, M-GCD shows a strong correlation between the star formation efficiency $\alpha_{\rm SF}$ and the mass loading factor η_0. However, this correlation cannot be found in the other two models. The variation among the posterior distributions of

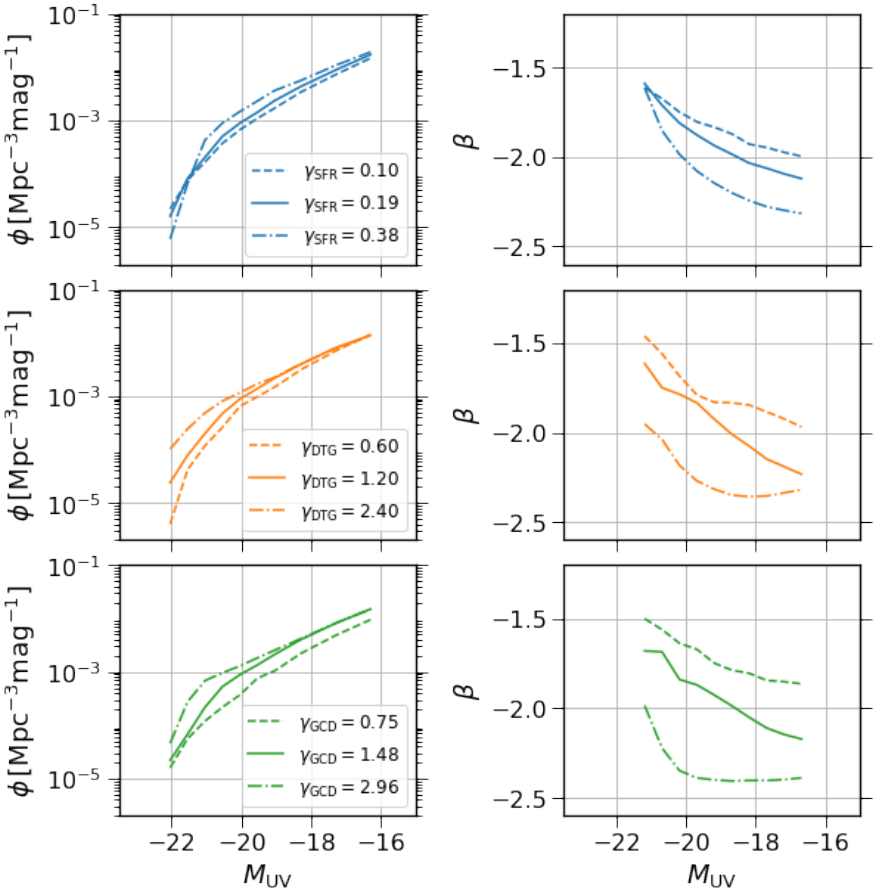


FIGURE 4.8: Effects of varying the galaxy property scaling of the dust relation $\gamma_{\rm SFR,\,DTG,\,GCD}$ on the dust-attenuated UV luminosity functions and colour magnitude relations. In each panel, solid lines correspond to the results of the best-fit models as shown in Figure 4.1. Dashed and dot dashed lines show the results of the model variants, in which $\gamma_{\rm SFR,\,DTG,\,GCD}$ is changed by a factor of two. From top to bottom, rows correspond to the dust attenuation model of M-SFR (Section 4.3.1), M-DTG (Section 4.3.2), M-GCD (Section 4.3.3) respectively.

the three models implies that these free parameters fit the observational data in a very complex way and the constraints on them depend on the assumptions used to model the dust attenuation.

By comparing the posterior distributions of the three dust models, I find similar correlations among the parameters of the supernova feedback, galaxy property scaling of the dust relation and reddening slope. I demonstrate this in Figure 4.6. It can be seen that the supernova energy coupling efficiency is positively and inversely correlated with $\gamma_{\rm SFR,\,DTG,\,GCD}$ and the reddening slope n, respectively. While the trends are the weakest for the M-GCD model, the correlation between $\gamma_{\rm SFR,\,DTG,\,GCD}$ and n is obvious for all the three models. Similar correlations are also found for the mass loading factor η_0. The

reader is referred to Figures 4.3, 4.4 and 4.5 for the two parameter marginalised distributions of all model parameters. The dependence between the galaxy formation and dust parameters is important, since it suggests that the observations can put constraints on intrinsic or dust-unattenauted galaxy properties despite the degrees of freedom in the dust models.

In order to understand the correlations mentioned above, I plot the intrinsic LFs and CMRs for the three best-fit models in Figure 4.1. They are shown as dashed lines. It can be seen that the LFs of the best-fit M-SFR is roughly a factor of two higher than for the other two models. I note that the intrinsic luminosity functions are more sensitive to feedback processes rather than the star formation law due to self-regulation (e.g. Schaye et al., 2010; Lagos et al., 2011). The supernova coupling efficiency of the best-fit M-SFR model is much smaller than the other two models, which is likely to be the main reason for the difference in the intrinsic LFs, irrespective of those star formation parameters. I examine the effect of supernova feedback in Figure 4.7. For the upper panel, I vary the mass loading factor η_0 at fixed supernova energy coupling efficiency ϵ_0 for the best-fit M-DTG model, and compare the resulting LFs with the best-fit results. The y-axis shows the ratio of the logarithmic intrinsic LFs. I find that the number density at fixed UV magnitude decreases with increasing η_0. Since the energy coupling efficiency puts an upper limit on the reheated mass (see Equation 2.9), the change in the LFs is smaller at larger η_0. The results of varying ϵ_0 at fixed η_0 are shown in the lower panel of Figure 4.7. While higher ϵ_0 decreases the LFs, the effect is found to be more significant at the bright end. Since the energy coupling efficiency is assumed to scale as a power law of the maximum circular velocity (see Equation 4.2), the efficiency can easily reach the maximum value of unity for small galaxies. The median $M_{\rm UV}$ - $V_{\rm max}$ relation of the best-fit M-DTG model indicates that the energy coupling efficiency becomes unity at an intrinsic magnitude $M_{\rm UV} \sim -18$ with $\epsilon_0 = 1.5$. Therefore, the major effect of increasing ϵ_0 is to allow more gas to be reheated in galaxies hosted by more massive halos. This explains why this parameter has larger impact at the bright end of the intrinsic LFs. Overall, the above discussion implies that supernova feedback plays an important role in regulating the intrinsic LFs.

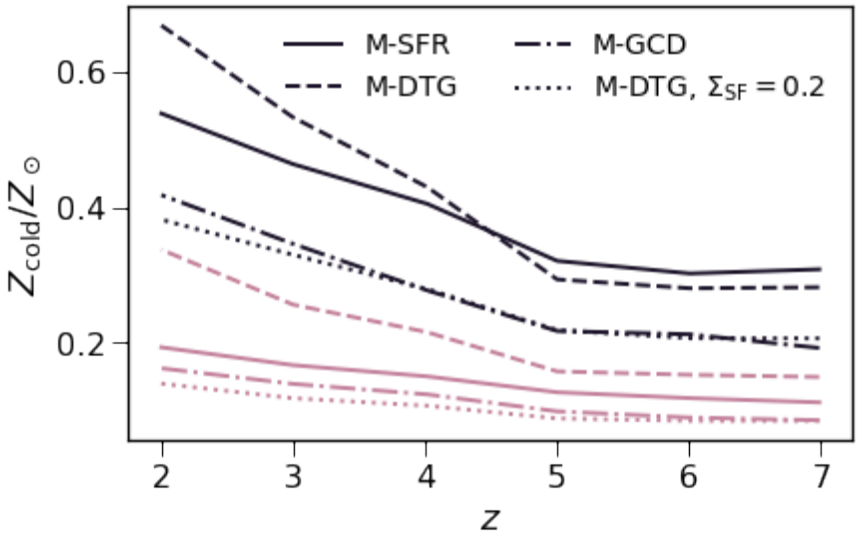
I next investigate the effect of the galaxy property scaling of the dust relation. Figure 4.8 shows the resulting dust-attenuated LFs and CMRs when $\gamma_{\rm SFR,DTG,GCD}$ is changed by a factor of two. It can be seen that the shape of both the LFs and CMRs are quite

sensitive to this parameter. Furthermore, since the dust optical depths are assumed to depend on different galaxy properties, this parameter changes the shape of the LFs and CMRs in different ways.

Combining the discussions of the supernova feedback parameters and $\gamma_{\rm SFR,DTG,GCD}$ above, I provide an explanation of the correlations between the model parameters in Figure 4.6. In my dust models, the effective UV optical depth is a function of the galaxy property scaling $\gamma_{\rm SFR,DTG,GCD}$ and the optical depth normalisations of both the ISM and BC. The galaxy property scaling has a direct impact on the shape of the dust-attenuated LFs and CMRs, and this single parameter is required to satisfy two shapes. Hence, the effective UV optical depth is very sensitive to $\gamma_{\rm SFR,DTG,GCD}$. The effective optical depth should be degenerate with the intrinsic UV LFs, which are primarily controlled by the supernova feedback parameters. These imply that both η_0 and ϵ_0 should be correlated with $\gamma_{\rm SFR,DTG,GCD}$. On the other hand, the observed UV continuum slope β depends on the reddening curve, which is assumed to be a power law of wavelength. Since there is a natural degeneracy between the power law slope and the normalisation, the reddening slope n should be degenerate with the effective UV optical depth, and therefore $\gamma_{\rm SFR,DTG,GCD}$. The dependence between the supernova feedback parameters and the reddening slope can be derived from the above two correlations. Figure 4.8 shows that the galaxy property scaling on the SFR changes the dust-attenuated LFs and CMRs differently from the other two models, which may explain why the best-fit M-SFR model requires much weaker supernova feedback. This also explains the shallower reddening slope required by the best-fit M-SFR model.

In addition, I contrast the best-fit models for M-DTG and M-GCD. Whilst their intrinsic LFs and CMRs are almost the same, a difference is found in the metallicity. Figure 4.9 illustrates the cold gas metallicity at two stellar mass bins as a function of redshift for all best-fit models. It is clear that the cold gas is more metal enriched in the best-fit M-DTG than in M-GCD. I identify that the normalisation of the critical mass $\Sigma_{\rm SF}$ is the primary driver for the difference since both best-fit models have similar parameters of supernova feedback. To confirm this, I run a model variant, setting $\Sigma_{\rm SF} = 0.2$, with other parameters being the same with the best-fit M-DTG. The resulting metallicity is also shown in Figure 4.9, which is similar to that of the best-fit M-GCD. This finding is unsurprising since the star formation law affects gas fraction and therefore metallicity (e.g. Schaye et al., 2010; Lagos et al., 2011). In addition, from Figure 4.9, it is worth

FIGURE 4.9: Redshift evolution of the mass metallicity relation.



The y-axis represents the cold gas metallicity, with $Z_\odot = 0.02$. Dark and light lines correspond to the metallicity at different stellar mass bins, $10^8\,\mathrm{M}_\odot < M_* < 10^{8.5}\,\mathrm{M}_\odot$ and $10^9\,\mathrm{M}_\odot < M_* < 10^{9.5}\,\mathrm{M}_\odot$ respectively. Solid, dashed and dot dashed lines show the best-fit results of M-SFR (Section 4.3.1), M-DTG (Section 4.3.2) and M-GCD (Section 4.3.3) respectively. The metallicity only depends on the galaxy formation parameters of these models, which are listed in Table 4.2. The dotted lines correspond to the results of a model variant for which the normalisation of critical mass is set to be $\Sigma_{\mathrm{SF}} = 0.2$, while other parameters remain the same with the best-fit M-DTG. This variant model illustrates that Σ_{SF} is a main driver of the cold gas metallicity.

noting that the metallicity evolves with redshift in my model, with higher metallicity at lower redshifts. This is expected due to the explicit redshift dependence on the mass loading factor, which is motivated by previous studies (Muratov et al., 2015; Hirschmann et al., 2016; Collacchioni et al., 2018).

4.7 Infrared excess to UV continuum slope relation

As demonstrated in previous sections, by simultaneously fitting the galaxy formation and dust models to the observed UV LFs and CMRs, We are able to obtain constraints on both the dust attenuation in the UV band and the reddening. This allows estimates of the infrared luminosity F_{IR} and therefore the infrared excess (IRX) using energy balance

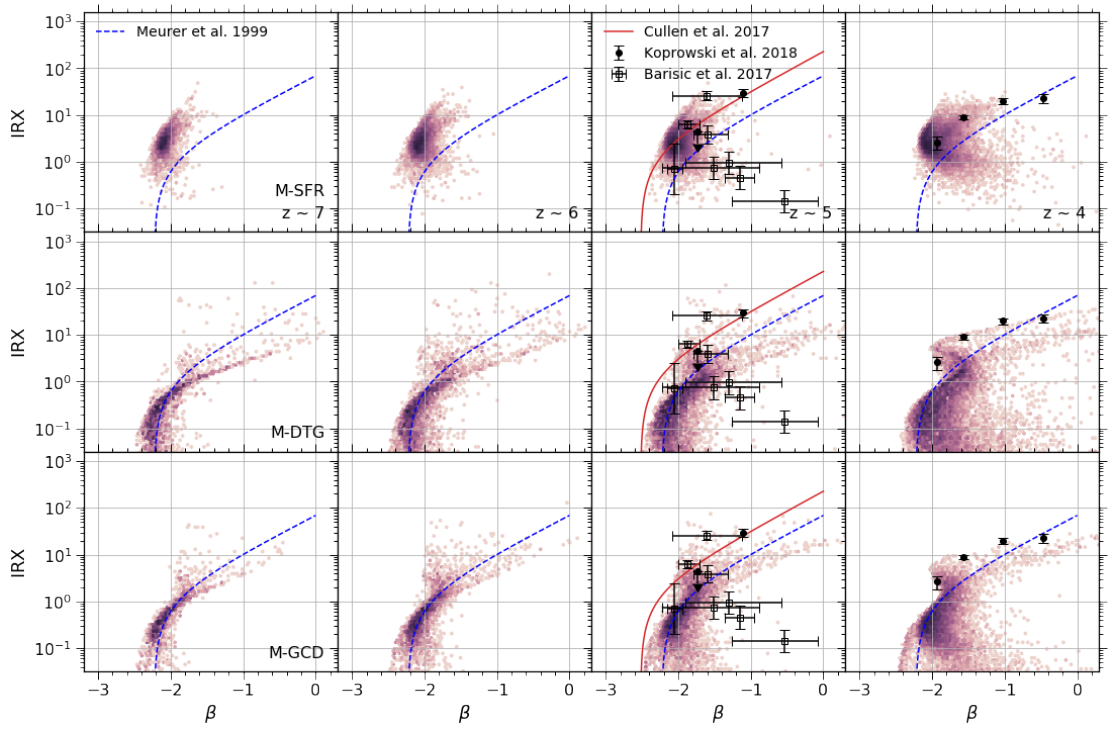


FIGURE 4.10: Predicted infrared excess (IRX) - UV continuum slope β relations. From top to bottom, rows show the results of the best-fit M-SFR (Section 4.3.1), M-DTG (Section 4.3.2) and M-GCD (Section 4.3.3) models. I only show model galaxies with stellar mass greater than $10^8\,M_\odot$. The dust optical depths in the three models are linked to the star formation rate (SFR), dust to gas ratio (DTG) and gas column density (GCD) respectively. The relations are represented by purple density plots. The best-fit parameters of these models can be seen in Table 4.2. IRX is computed by energy balance arguments. Columns show the results at different redshifts. Blue dashed lines are the widely used Meurer et al. (1999) relation. Red lines show the results from Cullen et al. (2017), which are based on the post-process of the FiBY hydrodynamic simulation (Johnson et al., 2013; Paardekooper et al., 2015). Black circles with error bars are stacking measurements of Lyman-break galaxies (LBGs) from Koprowski et al. (2018). Individual source measurements from Barisic et al. (2017) are shown as empty squares.

arguments, i.e.

$$F_{\rm IR} = \int_{912\text{Å}}^{\infty} \left(L_\lambda - L_\lambda^{\rm intrinsic}\right)\, d\lambda \tag{4.15}$$

$$F_{\rm UV} = \lambda L_\lambda\big|_{\lambda=1600\text{Å}} \tag{4.16}$$

$$\text{IRX} = \frac{F_{\rm IR}}{F_{\rm UV}} \tag{4.17}$$

I compute the IRX for galaxies in the best-fit models of M-SFR, M-DTG and M-GCD. The resulting IRX - β relations for galaxies with stellar mass greater than $10^8\,M_\odot$ are shown in Figure 4.10 with several observations for comparison. Taking into account intrinsic scatter in the relations, my results cover the observations from Koprowski et al.

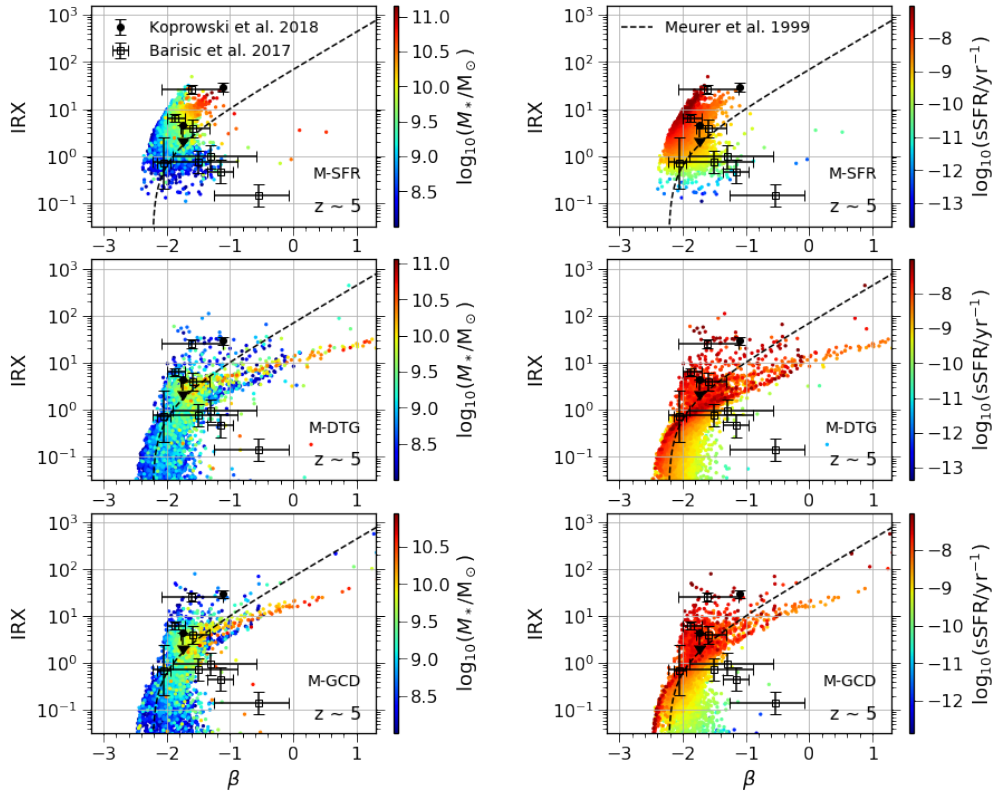


FIGURE 4.11: Predicted infrared excess (IRX) - UV continuum slope β relations as functions of stellar mass (left panels) and specific star formation rate (sSFR) (right panels) at $z \sim 5$. I only show model galaxies with stellar mass greater than $10^8\,\mathrm{M}_\odot$. From top to bottom, rows show the results of the best-fit M-SFR (Section 4.3.1), M-DTG (Section 4.3.2), M-GCD (Section 4.3.3). The dust optical depths in the three models are linked to the star formation rate (SFR), dust to gas ratio (DTG) and gas column density (GCD) respectively. Black dashed lines show the relation measured by Meurer et al. (1999). Black circles and empty squares with error bars are stacking and individual measurements from Koprowski et al. (2018) and Barisic et al. (2017) respectively.

(2018), who performed a stacking analysis of 4209 Lyman-break galaxies (LBGs) at $3 \lesssim z \lesssim 5$, and individual detection from Barisic et al. (2017). I also compare my predictions with the relations calibrated by Meurer et al. (1999) using local starburst galaxies. The Meurer et al. (1999) relation is frequently used to correct dust extinction in both observational and theoretical studies at high redshifts (e.g. Duncan et al., 2014; Bouwens et al., 2015; Mason et al., 2015; Liu et al., 2016; Harikane et al., 2018). It can be seen from Figure 4.10 that the best-fit M-SFR predicts higher IRX than the Meurer et al. (1999) relation at fixed β, while the other two best-fit models suggest lower IRX. Thus, my models indicate dust extinction that differs from the Meurer et al. (1999) relation, which implies that a direct application of the relation at high redshifts may lead to systematic errors on the dust corrections. I will discuss the resulting uncertainties

on estimations of the cosmic SFRD in Section 4.8.

4.7.1 Reddening slope

The best-fit models for M-SFR, M-DTG and M-GCD have quite different reddening slopes n, which can be read from Table 4.2. The best-fit M-SFR model has the shallowest slope of $n = -0.3$, while much steeper slopes are found for the best-fit M-DTG and M-GCD, with $n = -1.6$ and -1.3 respectively. This difference is directly reflected on the IRX - β plane. In Figure 4.10, the best-fit M-DTG and M-GCD show a shallower IRX - β relation at redder UV slope regime. Similar disagreements can also be found from other studies. For example, Cullen et al. (2017) post-processed the outputs of the FiBY hydrodynamic simulation (Johnson et al., 2013; Paardekooper et al., 2015). They propose a similar dust model to the present work, linking the dust optical depths to the logarithmic stellar mass. The free parameters in their model are adjusted to fit the observed LFs and CMRs from Rogers et al. (2014) at $z \sim 5$. They suggest $n = -0.55^{+0.25}_{-0.15}$. I plot their results as solid red lines in Figure 4.10, which is more consistent with the best-fit M-SFR than the other models. On the other hand, Mancini et al. (2016) also post-processed a hydrodynamic simulation by Maio et al. (2010), and coupled it with a dust evolution model. Their results reproduce the observed LFs of Bouwens et al. (2015) and CMRs of Bouwens et al. (2014) at $z \sim 5 - 8$ when using an SMC-like extinction curve. The slope of the SMC-like extinction curve is steeper, and is similar to those of the best-fit M-DTG and M-GCD. Since all the models can well reproduce observed LFs and CMRs, I cannot draw any firm conclusions on the reddening slope. Instead, I treat this as systematic uncertainties arising due to different assumptions in the dust models.

4.7.2 Intrinsic scatter

At $z \gtrsim 3$, observations show considerable scatter in the IRX - β plane (e.g. Capak et al., 2015; Álvarez-Márquez et al., 2016; Bouwens et al., 2016; Barisic et al., 2017; Fudamoto et al., 2017; Koprowski et al., 2018), which might be explained by the large intrinsic scatter in the predicted relations. Hence, it is instructive to examine the main drivers of the scatter. I first notice that from Figure 4.10, low IRX galaxies vanish in the best-fit M-SFR model. This is due to the nature of my star formation prescription

(see Section 2.3). The SFR of galaxies whose cold gas mass is below the critical mass is zero. Accordingly, in the M-SFR model, the dust optical depths of these galaxies are also zero, which results in the disappearance of the IRX. This unrealistic feature shows the limitations of this model.

In left and right panels of Figure 4.11, I illustrate the IRX - β relations at $z \sim 5$ as functions of stellar mass and specific star formation rate (sSFR) respectively. The relations at other redshifts show similar trends. For the stellar mass case, we see that massive galaxies form a tight correlation between IRX and β in the high IRX and red β regions. The trend that more massive galaxies have higher IRX is also observed by Álvarez-Márquez et al. (2016) and Fudamoto et al. (2017). However, I also find several larger stellar mass galaxies which have lower IRX and redder β. They might explain some of the outliers individually detected by Barisic et al. (2017), as shown in Figure 4.11. On the other hand, the right panels show that the scatter of the IRX - β relation is tightly correlated with sSFR. At fixed IRX, redder galaxies typically have lower sSFR. This trend is consistent with other theoretical studies (Popping et al., 2017b; Safarzadeh et al., 2017; Narayanan et al., 2018; Cousin et al., 2019). In addition, it is worth noting that although the dust optical depths are related to different galaxy properties for the three dust models, I find similar dependence of the scatter in the IRX - β plane on both stellar mass and sSFR.

4.8 Cosmic star formation density

Dust corrections are typically required for the conversion between the UV luminosity and the SFR. As mentioned, in high redshift observational studies of SFR, the Meurer et al. (1999) relation is widely used, though it is calibrated against local galaxies. The previous section has shown that the dust extinction predicted by my models, which reproduce both LFs and CMRs at $z \sim 4 - 7$, is rather different from the Meurer et al. (1999) relation. In principle, I could derive similar relations based on my results to be used by other studies to perform the dust corrections. However, by using such relations, I should be able to recover the SFR functions of my models given the LFs. Therefore, I directly present the predicted SFRs. Furthermore, the difference among the three dust models allows us to estimate the systematic uncertainties in the observed SFRs.

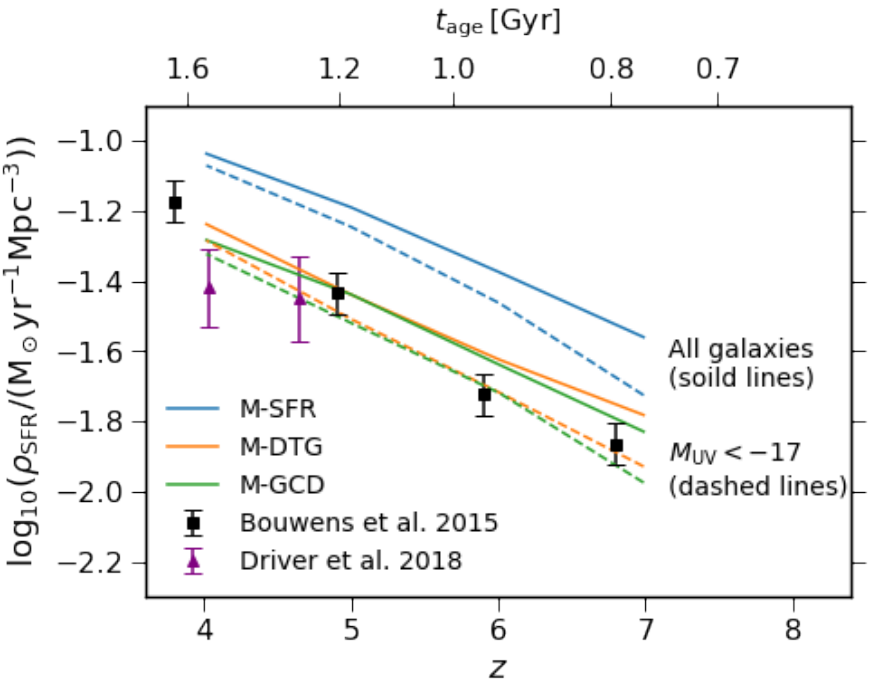


FIGURE 4.12: Predicted cosmic star formation rate density (SFRD) at $z \sim 4-7$. Blue, orange and green lines are estimated from the best-fit M-SFR (Section 4.3.1), M-DTG (Section 4.3.2), M-GCD (Section 4.3.3) respectively. The dust optical depths in the three models are linked to the star formation rate (SFR), dust to gas ratio (DTG) and gas column density (GCD). Solid lines are the SFRD of all model galaxies, while the results with a magnitude cut $M_{UV} < -17$ are shown as dashed lines. Black data points are observations from Bouwens et al. (2015). Their dust corrections are derived by the colour magnitude relations (CMRs) of Bouwens et al. (2014) and the Meurer et al. (1999) relation. Purple trangles with error bars show the measurements of Driver et al. (2018) using the energy balance SED-fitting code MAGPHYS (da Cunha et al., 2008).

Figure 4.12 illustrates the predicted cosmic star formation rate density (SFRD) for the best-fit models of M-SFR, M-DTG and M-GCD. Their values are listed in Table 4.3. I compare the results with Bouwens et al. (2015), whose estimations are based on the CMRs of Bouwens et al. (2014) and the Meurer et al. (1999) relation. Since the results of Bouwens et al. (2015) and my models use the same observational information, the comparison between them quantifies the systematic errors of using the Meurer et al. (1999) relation with respect to correcting the dust extinction. I note that all my models suggest bluer intrinsic UV continuum slopes than the one used in Meurer et al. (1999) as shown in Figure 4.1. Figure 4.10 also illustrates that the best-fit results of M-DTG and M-GCD have shallower IRX - β relations. Thus, compared with the Meurer et al. (1999) relation, the dust extinction in these two models is stronger for bluer galaxies but weaker for redder galaxies. On the other hand, the dust attenuation is stronger for all galaxies in the best-fit M-SFR. It can be seen from Figure 4.12 that the cosmic SFRD of the

TABLE 4.3: Tabular data of predicted cosmic star formation rate density (SFRD) for the three different dust models.

z	All galaxies			$M_{\rm UV} < -17$		
	M-SFR	M-DTG	M-GCD	M-SFR	M-DTG	M-GCD
4	-1.04	-1.24	-1.28	-1.07	-1.29	-1.32
5	-1.19	-1.44	-1.44	-1.25	-1.51	-1.52
6	-1.38	-1.62	-1.64	-1.46	-1.72	-1.72
7	-1.56	-1.78	-1.83	-1.73	-1.93	-1.97

These values are in a unit of $\log_{10}(\rho_{\rm SFR}/(\,{\rm M}_\odot{\rm yr}^{-1}{\rm Mpc}^{-3}))$.

best-fit M-DTG and M-GCD are consistent with those of Bouwens et al. (2015), while the results of the best-fit M-SFR is roughly a factor of two higher. I also compare my results with Driver et al. (2018). Their dust-corrected SFRs are derived from the energy balance SED-fitting code MAGPHYS (da Cunha et al., 2008). Better consistency is found between their measurements and my best-fit models of M-DTG and M-GCD, given the size of the error bars on those points. Overall, Figure 4.12 suggests that uncertainty in the dust relations introduces at least a factor of two systematic error into the inferred cosmic SFRD at $z \gtrsim 6$.

4.9 Summary

This work investigates the IRX - β relation and cosmic SFRD at $z \sim 4-7$ by combining the the MERAXES semi-analytic galaxy formation model (Mutch et al., 2016; Qin et al., 2017) and the Charlot & Fall (2000) dust attenuation model. The supernova feedback model of MERAXES is updated using results from previous studies (Muratov et al., 2015; Hirschmann et al., 2016; Cora et al., 2018), which aim to reproduce the redshift evolution of the mass metallicity relation. I introduce three different parametrisations of the dust optical depths, which are related to the star formation rate (M-SFR), dust-to-gas ratio (M-DTG) and gas column density (M-GCD) respectively. These lead to five free parameters in each dust model in addition to those in MERAXES.

The determinations on not only the dust parameters but also the MERAXES free parameters constitute the primary part of this work. For galaxy formation parameters, I focus on the star formation efficiency, critical mass, mass loading factor and supernova coupling efficiency. I adopt a Bayesian approach, calibrating these parameters against

the UV luminosity functions (LFs) of Bouwens et al. (2015) and colour magnitude relations (CMRs) of Bouwens et al. (2014) at $z \sim 4-7$. The posterior distribution of these parameters is estimated using multimodal nested sampling (Feroz et al., 2009). I find that these observations can be fit extremely well by all the three dust models. However, the preferred parameter ranges are quite different among the three dust models. My analysis indicates that the combination of the LFs and CMRs can put strong constraints on a given dust attenuation model, since the model is required to reproduce the shape of both observations. The differences in my results are due to the different assumptions of the dust models, which results in different relations between UV dust attenuation and intrinsic UV magnitude.

I then demonstrate the predictions of my calibration results. Using energy balance arguments, I estimate the IRX for each model galaxy. I find that the predicted IRX - β relations are quite different from the Meurer et al. (1999) relation, and contain large intrinsic scatter, which might explain the current discrepancy among several high redshift observations (e.g. Capak et al., 2015; Álvarez-Márquez et al., 2016; Bouwens et al., 2016; Barisic et al., 2017; Fudamoto et al., 2017; Koprowski et al., 2018). I also confirm the correlation between the intrinsic scatter and sSFR. This finding is consistent with other theoretical studies (Popping et al., 2017b; Safarzadeh et al., 2017; Narayanan et al., 2018; Cousin et al., 2019). Secondly, I present model predictions for the cosmic SFRD, and compare these with the observations of Bouwens et al. (2015) and Driver et al. (2018). The difference among the three dust models implies at least a factor of two systematic uncertainty in the observed SFRD when corrected using the Meurer IRX - β relation.

This work has simultaneously constrained the free parameters of a semi-analytic galaxy formation model and additional dust parameters using observations of UV properties. Within a Bayesian framework, my approach establishes a more direct connection between the model and observations despite the complexity. This approach is particularly useful for studies at high redshifts where UV properties are the most robust observables. This work could be further improved by explicitly modelling the dust evolution (e.g Mancini et al., 2016; Popping et al., 2017a; Dayal & Ferrara, 2018), which might reduce the systematic uncertainties due to different assumptions in the dust models. Additional free parameters (e.g. the time scale of dust growth) in such dust evolution models could also be constrained using my methodology.

Chapter 5

Extending mass resolution of N-body simulations

5.1 Introduction

In the previous two chapters, I demonstrated the applications of the MERAXES semi-analytic model in studying high redshift galaxies, while in this chapter, I propose a method to improve the predictions of cosmic reionisation from semi-analytic models. Simulating the epoch of reionisation is extremely challenging, with different techniques developed to study different aspects of the problem. For example, high resolution hydrodynamical simulations (e.g. Wise et al., 2012; Johnson et al., 2013; Ceverino et al., 2017; Rosdahl et al., 2018) can resolve the faintest galaxies with detailed spatial information on the interstellar media (ISM). These faint sources are found to have non-negligible contributions to reionisation (Wise et al., 2014; Katz et al., 2020). However, these simulations are limited to a small volume ($\lesssim 10^3\,h^{-3}\mathrm{Mpc}^3$). At the other extreme, Iliev et al. (2014) presented a study in a $425\,h^{-1}$Mpc box, and pointed out that at least a $\sim 100\,h^{-1}$Mpc box is required for the convergence of reionisation histories. To achieve this volume, they used a mass-to-light ratio to assign ionising sources to dark matter halos in a large N-body simulation. Other studies use semi-numerical calculations of reionisation to simulate large volumes (e.g Greig & Mesinger, 2015; Hassan et al., 2016; Park et al., 2019). A disadvantage of these approaches is the absence of a physically

motivated galaxy formation model. Whilst large volumes have been achieved by several hydrodynamical simulations (e.g. Feng et al., 2016; Pillepich et al., 2018), they are extremely computationally expensive and cannot resolve the faintest sources.

Semi-analytic models have great potential to satisfy the demand of both high mass resolution and large volume. The mass resolution and simulation volume of semi-analytic modes are determined by the input N-body simulations. For the purpose of modelling cosmic reionisation, very large N-body simulations are needed. In this chapter, I present a methodology to extend the mass resolution of N-body simulations using Monte Carlo algorithms. The resulting halo catalogues are designed for both semi-analytic models and reionisation calculations.

This chapter is organised as follows. Section 5.2 introduces the N-body simulations used in this study. The methods to augment N-body merger trees, assign halo positions, and sample halo spin parameters are described in Sections 5.3, 5.4, and 5.5 respectively. I apply the extended halo catalogues to the MERAXES semi-analytic model in Section 5.6. Finally, this work is summarised in Section 5.7.

5.2 The Genesis N-body simulations

I utilise L105N2048 and L35N2650 from the Genesis N-body simulations (Elahi et al., in preparation). The methodology of extending the halo mass resolution is applied to L105N2048, which is a $105\,h^{-1}$Mpc box, containing 2048^3 particles, with $m_{\rm p} = 1.17 \times 10^7\ h^{-1}\,{\rm M}_\odot$. L35N2650 is used to calibrate and verify the results. It contains 2650^3 particles in a $35\,h^{-1}$Mpc box. The particle mass is $m_{\rm p} = 2.00 \times 10^5\,h^{-1}\,{\rm M}_\odot$. All the simulations are run using GADGET-2 (Springel, 2005). Halos in the simulations are identified using VELOCIRAPTOR (Elahi et al., 2019c,a), which is a six-dimensional friends-of-friends phase space halo finder. Merger trees are constructed using TREEFROG (Elahi et al., 2019d,b). Table 5.3 provides a summary of the N-body halo catalogues used in this work. Throughout the paper, I adopt the mass obtained by summing all particles in a friends-of-friends group as halo mass. Throughout this chapter, I adopt the same cosmology as the Genesis simulations, with $h = 0.6751$, $\Omega_{\rm m} = 0.3121$, $\Omega_{\rm b} = 0.0491$, $\Omega_\Lambda = 0.6879$, $\sigma_8 = 0.8150$, $n_{\rm s} = 0.9653$ (fourth column in Table 4 of Planck Collaboration et al., 2016).

TABLE 5.1: Parameters of the Monte Carlo tree algorithm.

Symbol	Parkinson et al. (2008)	This work
G_0	0.57	1.0
γ_1	0.38	0.2
γ_2	-0.01	-0.4

5.3 Augmenting N-body merger trees

My approach to augment N-body merger trees mainly follows Benson et al. (2016). The basic idea is to generate Monte Carlo merger trees with the desired mass resolution and compare these with an N-body merger tree in the mass range where the simulation is fully reliable. If both trees are similar, as determined by several criteria (described below), Monte Carlo halos with mass below the simulation resolution are attached to the N-body merger tree. This results in a hybrid structure, containing both Monte Carlo and N-body halos, but with the same mass resolution as the Monte Carlo tree.

5.3.1 Generating Monte Carlo trees

I adopt the Parkinson et al. (2008) algorithm to generate Monte Carlo merger trees. The algorithm is based on binary splits in small internal time steps. It employs the conditional mass function [1] derived from the Extended Press Schechter (EPS) theory (Bower, 1991; Bond et al., 1991; Lacey & Cole, 1993) with an additional parameterisation to take into account the difference between the EPS theory and N-body simulations. The conditional mass function is expressed as

$$f(M_1, z_1|M_2, z_2) = G_0 \left(\frac{\sigma_1}{\sigma_2}\right)^{\gamma_1} \left(\frac{\delta_2}{\sigma_2}\right)^{\gamma_2} f_{\rm EPS}(M_1, z_1|M_2, z_2), \tag{5.1}$$

where $f_{\rm EPS}(M_1, z_1|M_2, z_2)$ is the conditional mass function given by the EPS theory. I denote $\sigma_1 = \sigma(M_1)$ and $\sigma_2 = \sigma(M_2)$, which are the mass variance of the matter density field linearly extrapolated to $z = 0$ and smoothed by a spherical tophat filters at M_1 and M_2. The density contrast is defined by $\delta_2 = 1.686/D(z_2)$, where $D(z)$ is the linear growth factor. The free parameters are G_0, γ_1 and γ_2. Parkinson et al. (2008) calibrated these free parameters against the Millennium simulation (Springel et al., 2005) in the mass range between $10^{12}\,h^{-1}\,{\rm M}_\odot$ and $10^{15}\,h^{-1}\,{\rm M}_\odot$ and from $z = 0$ to $z = 4$. However, in

[1] The conditional mass function discussed here is defined by the mass fraction distribution (M_1/M_2) as a function of progenitor mass M_1 given the descendant mass M_2.

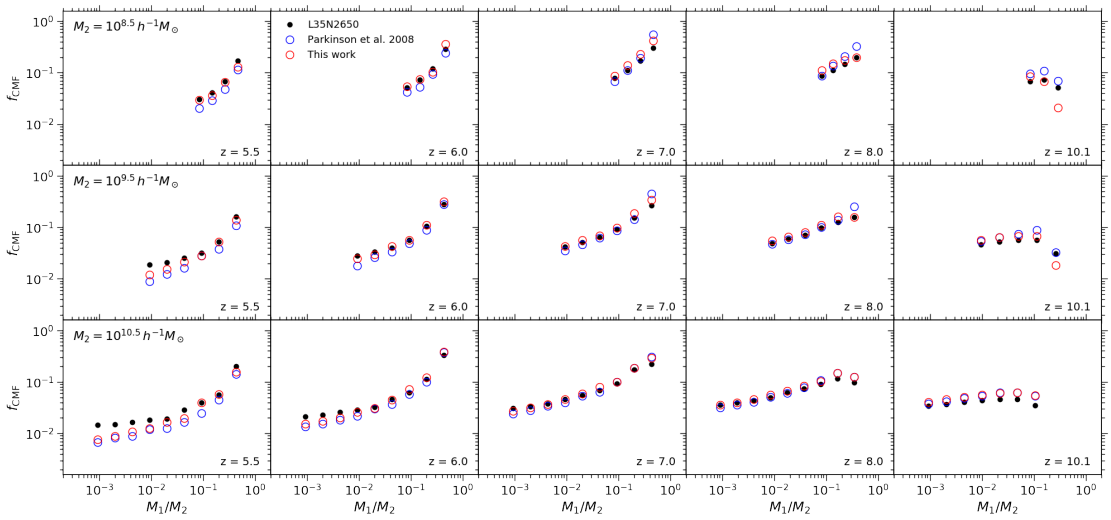


FIGURE 5.1: Fitting results of the calibration for the Parkinson et al. (2008) algorithm. The conditional mass functions are defined by $df_{\rm CMF}/d\ln M_1$. Black dots are the fitting data, which are estimated using L35N2650. Red and blue empty circles are the results corresponding to the best-fit parameters obtained in this work and those used by Parkinson et al. (2008) respectively. The values of the parameters are listed in Table 5.1.

this work, I am interested in growing halos at $z \geq 5$, and require that the mass resolution of the merger trees reaches the atomic cooling threshold ($\sim 10^7 - 10^8\,h^{-1}M_\odot$). Therefore, I recalibrate the parameters against our simulations, which also accounts for updated cosmology.

Following Parkinson et al. (2008), the cost function of the calibration is given by

$$\mathcal{C}(G_0, \gamma_1, \gamma_2) = \sum \left[\log_{10} f_{\rm NS} - \log_{10} f_{\rm MC}\right]^2 , \tag{5.2}$$

where $f_{\rm NS}$ and $f_{\rm MC}$ are the conditional mass functions of the N-body and Monte Carlo merger trees respectively. I estimate $\log_{10} f_{\rm NS}$ from L35N2650 and $\log_{10} f_{\rm MC}$ using samples of 300 Monte Carlo merger trees for each descendant mass M_2. The fitting points calculated from the simulation are shown as black dots in Figure 5.1. I employ the particle swarm optimisation (Shi & Eberhart, 1998) to minimise the cost function. The best-fit parameters are accepted if they do not change for 100 iterations. Their values are given in Table 5.1, and the fitting results are illustrated in Figure 5.1. My best-fit parameters improve the cost function by $\Delta\mathcal{C} \approx -0.6$, compared with Parkinson et al. (2008).

5.3.2 Augmentation algorithm

The most important and difficult component of the augmentation is to decide whether a Monte Carlo tree is similar to an N-body tree. Instead of comparing entire trees, Benson et al. (2016) decompose an N-body merger tree into many sub-branches, and match only one sub-branch every time with Monte Carlo realisations. A sub-branch is comprised of one descendant halo and all halos that directly merge into it. Hereafter, I refer to this structure as a "simple branch".

I denote the mass of each progenitor in an N-body simple branch as M_1, M_2, ..., M_n with $M_1 > M_2 > ... > M_n$, where n is the number of the progenitors, and let $n_{\rm cut}$ be the number of the progenitors whose mass is above a threshold $M_{\rm cut}$. I use primed symbols for the same quantities of Monte Carlo trees. Benson et al. (2016) match N-body and Monte Carlo simple branches using:

(a) $n' \geq n_{\rm cut}$,

(b) for $i = 1, 2, ..., n_{\rm cut}$, $|M_i - M'_i| < \xi M_i$,

(c) for $i = n_{\rm cut} + 1, n_{\rm cut} + 2, ..., n'$, $M'_i < M_{\rm cut}$,

where ξ is a free parameter and controls the mass precision of the match. Once a match is found, N-body progenitors at $M_{\rm halo} < M_{\rm cut}$ are replaced by Monte Carlo halos in the same mass range. In the resulting hybrid structure, the descendant halo and progenitors with mass above $M_{\rm cut}$ are from the original simple branch, while progenitors with mass below $M_{\rm cut}$ are additional Monte Carlo halos from the match.

In practice, relaxing the three matching criteria (a), (b) and (c) is necessary, since there is often no match even for large numbers of Monte Carlo realisations. Benson et al. (2016) increase ξ by a factor of $1 + \epsilon_{\rm mass}$ after $N_{\rm mass}^{\rm limit}$ rejections. However, this only impacts the second condition. I have also found many cases where the first and third conditions are never satisfied. This problem was not reported in Benson et al. (2016), and the reason might be that the mass range investigated in this work is much lower than in that paper. To address this issue, I increase $M_{\rm cut}$ by a factor of $1 + \epsilon_{\rm cut}$ after $N_{\rm cut}^{\rm limit}$ rejections. I do not allow $M_{\rm cut}$ to be greater than either a maximum mass cut $M_{\rm cut}^{\rm max}$ or the mass of the most massive progenitor. Furthermore, a maximum number of trials

$N_{\rm tot}^{\rm limit}$ is employed. Once this number of trials is reached, the algorithm is terminated and returns the input simple branch, with all progenitors below the minimum mass cut $M_{\rm cut}^{\rm min}$ removed. This treatment may remove some N-body halos without augmentation of Monte Carlo halos. However, in practice, I find that this situation occurs at a rate that is always smaller than 0.06% for a given snapshot.

N-body merger trees have a special feature that should be taken into account in the comparison with Monte Carlo merger trees. When the halo finder fails to identify the descendant of an N-body halo in the next snapshot, it may try to search for the descendant in later snapshots. Hence, progenitors in an N-body simple branch are not always from the adjacent snapshot. However, this situation never happens for Monte Carlo merger trees. I follow Benson et al. (2016) to resolve the issue. In order to make the trees comparable, for a given N-body simple branch, I manually set all progenitors to be located in the previous snapshot relative to their descendant, and keep their mass unchanged (except for the most massive progenitor, whose mass is interpolated with time).

N-body merger trees typically contain subhalos, which is an additional feature that Monte Carlo merger trees do not have. Following Benson et al. (2016), I do not consider subhalos in the tree augmentation. Accordingly, I reconstruct a merger trees that only consists of host halos from an original N-body trees. The reconstruction proceeds forward with time. If the descendant of an N-body halo is a subhalo, I redirect it to the host of the subhalo. I neglect the descendant of a subhalo when building the host halo merger trees.

In reconstructed N-body merger trees, I have found many massive halos ($M_{\rm halo} \gtrsim 10^{10}\,h^{-1}\,{\rm M}_\odot$) that have no progenitors. In the original trees, these halos only have one subhalo progenitor whose host mergers into a different target. When augmenting such halos, criteria (a) and (b) are automatically satisfied. However, I find that forcing criterion (c) overestimates the conditional mass function at $M_{\rm halo} < M_{\rm cut}$. Based on several experiments, I suggest the following modification, which can lead to more consistent conditional mass functions

(c′) if $n > 0$, for $i = n_{\rm cut}+1, n_{\rm cut}+2, ..., n'$, $M_i' < M_{\rm cut}$, otherwise for $i = 1, 2, ..., n'$, $M_i' < M_{\rm cut}^{\rm max}$.

Overall, given a simple branch in an N-body merger tree, my augmentation algorithm proceeds as follows:

1. Set $N_{\rm cut}^{\rm trial} = 0$, $N_{\rm mass}^{\rm trial} = 0$, $N_{\rm tot}^{\rm trial} = 0$, $\xi = \xi_0$, $M_{\rm cut} = M_{\rm cut}^{\rm min}$.
2. Whenever a progenitor is at a non-adjacent snapshot of its descendant halo, put it to one previous snapshot of the descendant. If the progenitor is the most massive, interpolate its mass with time.
3. Generate a Monte Carlo simple branch using the same configuration as the given N-body branch. Increase $N_{\rm tot}^{\rm trial}$ by 1.
4. Compare the N-body and Monte Carlo simple branches using criteria (a), (b) and (c′). If all of three criteria are satisfied, go to step 7, otherwise, increase the corresponding counters:
   - If criteria (a) or (c′) are false, increase $N_{\rm cut}^{\rm trial}$ by 1.
   - If criterion (b) is false, increase $N_{\rm mass}^{\rm trial}$ by 1.
5. Relaxing the criteria when certain number of rejections is reached:
   - If $N_{\rm cut}^{\rm trial} = N_{\rm cut}^{\rm limit}$, set $N_{\rm cut}^{\rm trial} = 0$ and increase $M_{\rm cut}$ by a factor of $1 + \epsilon_{\rm cut}$. If $M_{\rm cut}$ is greater than $M_{\rm cut}^{\rm max}$ or the mass of the most massive progenitors of the given simple branch, set it to be the minimum of these two values.
   - If $N_{\rm mass}^{\rm trial} = N_{\rm mass}^{\rm limit}$, set $N_{\rm mass}^{\rm trial} = 0$ and increase ξ by a factor of $1 + \epsilon_{\rm mass}$.
6. Terminate the algorithm if $N_{\rm tot}^{\rm trial} = N_{\rm tot}^{\rm limit}$, otherwise go to step 3.
7. Replace progenitors with mass below $M_{\rm cut}$ at the N-body simple branch with Monte Carlo halos in the same mass range.

I apply the augmentation algorithm to every halo in the N-body simulation backward with time, and grow new Monte Carlo halos using the Parkinson et al. (2008) algorithm. A schematic diagram of the augmentation can be found in Figure 5.2.

Free parameters in the algorithm are summarised in Table 5.2. Ideally, if the conditional mass functions of Monte Carlo merger trees are consistent with the N-body simulations, these parameters should primarily affect numerical efficiency and be insensitive to the results. However, as demonstrated in Figure 5.1, even with recalibrated parameters,

TABLE 5.2: Parameters of the tree augmentation algorithm.

Symbol	Value
ξ_0	0.2
$\epsilon_{\rm mass}$	0.2
$N_{\rm mass}^{\rm limit}$	50
$M_{\rm cut}^{\rm min}$	$100\,m_{\rm p}$ [a]
$M_{\rm cut}^{\rm max}$	$2500\,m_{\rm p}$ [a]
$\epsilon_{\rm cut}$	2.0
$N_{\rm cut}^{\rm limit}$	5
$N_{\rm tot}^{\rm trail}$	1000

[a] For L105N2048, $m_{\rm p} = 1.17 \times 10^7\,h^{-1}\,{\rm M}_\odot$, which is the particle mass of the simulation.

TABLE 5.3: Information on halo catalogues used in this work.

Name	Type	$l_{\rm box}$ $[h^{-1}{\rm Mpc}]$	m_p $[h^{-1}M_\odot]$	$m_{\rm res}$ $[h^{-1}M_\odot]$
L35N2650	N-body simulation	35	2.00×10^5	-
L105N2048	N-body simulation	105	1.17×10^7	-
L105E5	Hybrid	105	-	1.4×10^8
L105E10	Hybrid	105	-	5.7×10^7
L105E15	Hybrid	105	-	3.2×10^7

The mass resolutions of L105E5, L105E10 and L105E15 correspond to the atomic cooling threshold at $z = 5$, 10 and 15 respectively.

the Parkinson et al. (2008) algorithm is unable to reproduce all parts of the conditional mass functions of the N-body merger trees, particularly at the lower mass end and higher redshifts. For this reason, I find that the choice of the algorithm parameters impacts the resulting conditional mass functions. The values listed in Table 5.2 are chosen based on several experiments in order to obtain better consistency with the N-body simulations.

To summarise, my augmentation algorithm differs from Benson et al. (2016) by changing the mass cut $M_{\rm cut}$ dynamically (and introducing the maximum mass cut $M_{\rm cut}^{\rm max}$). When applying the approach of Benson et al. (2016), the result contains only Monte Carlo halos at $M_{\rm halo} < M_{\rm cut}$ and only N-body halos at $M_{\rm halo} > M_{\rm cut}$. In my approach, $M_{\rm cut}$ is not a constant. The minimum and maximum mass cuts become the dividing lines of N-body and Monte Carlo halos. At the mass range in between, halo types are mixed. This modification averages the difference between the merger trees extracted from N-body simulations and those generated by the Monte Carlo algorithm.

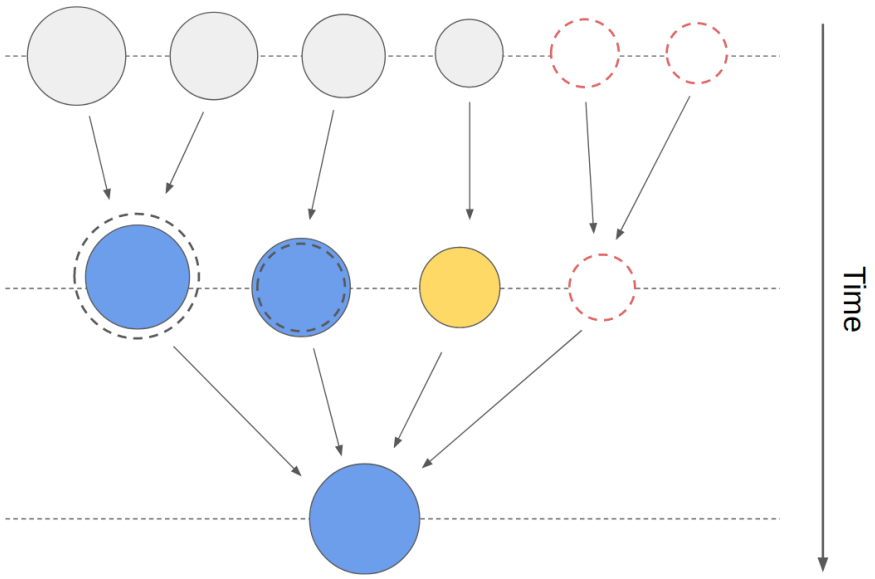


FIGURE 5.2: Schematic diagram of augmenting N-body halo merge trees. Solid and dashed circles represent N-body and Monte Carlo halos respectively, with size proportional to halo mass. The blue and yellow circles form an N-body simple branch (defined in Section 5.3.2), which is compared with a Monte Carlo tree. The algorithm removes halos with mass below $M_{\rm cut}$, corresponding to the yellow circle. The progenitors of removed halos will not be taken into account in the next step. Red dashed circles represent Monte Carlo halos that are added to the N-body simulation. The Monte Carlo halos on the top are grown from its descendant using the Parkinson et al. (2008) algorithm.

5.3.3 Fixing original subhalo trees

In N-body simulations, secondary progenitors may still be self-bounded for a certain period after a merger. Such objects are known as subhalos. They are invisible to the augmentation algorithm introduced in the previous section. A valid subhalo, by definition, must have a progenitor, a descendant and a halo to host it. Each of them can be removed during the augmentation if its mass is below $M_{\rm cut}$. To fix the problem, I first remove all subhalos that lose their host. Secondly, if a subhalo merges into a removed halo, I redirect it to the descendant of its host halo. For subhalos with no progenitors, we can prevent the applied semi-analytic model from seeding a galaxy in them.

5.3.4 Identifying the complete halo population

A complete halo population cannot be obtained by applying the augmentation algorithm introduced in Section 5.3. This is because there are halo merger trees that are never

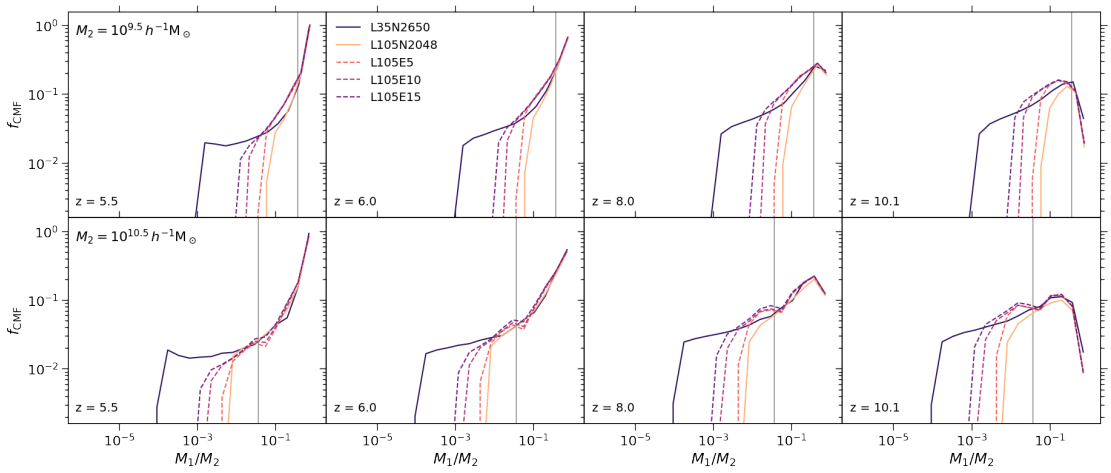


FIGURE 5.3: Comparisons of the conditional functions, defined by $df_{\mathrm{CMF}}/d\ln M_1$, of N-body and augmented merger trees. Solid lines are the results derived using L35N2650 and L105N2048. The information on these two N-body simulations can be found in Table 5.3. Dashed lines are based on augmented halo merger trees, which are obtained by applying the algorithm described in Section 5.3 to L105N2048. Deeper colours correspond to higher mass resolution. The grey vertical lines show the minimum mass cut of the augmentation algorithm.

resolved by the N-body simulation. Such merger trees do not interact with any halos in the simulation, and therefore cannot be identified during the Monte Carlo augmentation.

As a specific example in this work, I apply the augmentation algorithm at $z = 5$, adding Monte Carlo halos to the N-body merger trees backwards in time. However, at $z = 5$, the algorithm does not add new halos that are not resolved (between $M_{\mathrm{cut}}^{\mathrm{min}}$ and M_{res}). In addition, we also miss progenitors of such unresolved halos in earlier snapshots, resulting in an incomplete halo population. To fix this problem, I create an additional halo catalogue at $z = 5$, using masses and numbers drawn from the halo mass function of our N-body simulations. I use interpolation of a histogram instead of a fitting model for the halo mass function. I then generate trees for these halos using the Parkinson et al. (2008) algorithm. Hereafter, Monte Carlo halos generated by the augmentation algorithm are labelled as MC-I, while those in the additional catalogue are referred to as MC-II.

5.3.5 Applying to N-body simulations

I apply the approach introduced in the proceeding sections to augment the N-body merger trees of L105N2048 from $z = 5$ to $z = 20$. I choose three levels of mass resolution: $M_{\mathrm{res}} = 1.4 \times 10^8$, 5.7×10^7 and $3.2 \times 10^7\,h^{-1}\,\mathrm{M}_\odot$, corresponding to the atomic cooling

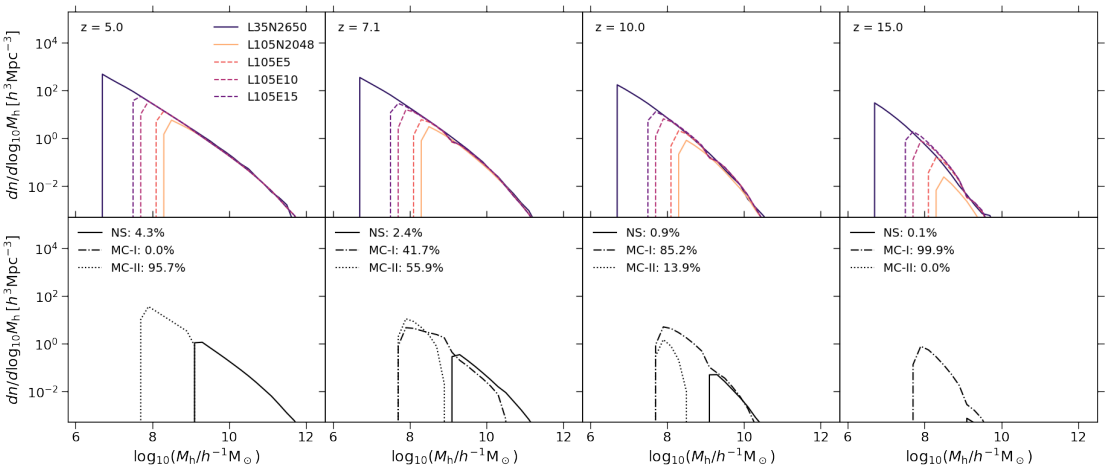


FIGURE 5.4: Halo mass functions of extended halo catalogues. Upper panels: comparisons of the halo mass functions of N-body and extended halo catalogues. Solid lines are estimated from the N-body simulations, using L35N2650 and L105N2048. Their information can be found in Table 5.3. Dashed lines are based on extended halo catalogues, which are obtained by applying the algorithm described in Section 5.3.2 to L105N2048. The mass resolutions of L105E5, L105E10 and L105E15 correspond to the atomic cooling thresholds at $z = 5$, 10 and 15 respectively. Deeper colours correspond to higher mass resolution. Bottom panels: halo mass functions of N-body, MC-I and MC-II halos from L105E10. Their mass fractions are labelled in the top left corners. See Section 5.3.4 for the definition of MC-I and MC-II halos.

threshold at $z = 5$, $z = 10$ and $z = 15$ respectively. These three extended halo catalogues are labelled as L105E5, L105E10 and L105E15. Their information is summarised in Table 5.3.

To test the results, I compare the conditional mass functions of augmented merger trees with our L35N2650 high resolution simulation in Figure 5.3. Upper and lower panels correspond to different descendant halo mass bins. The conditional mass functions (CMFs) of extended trees are shown as dashed lines, which broadly agree with L35N2650. Several discrepancies, e.g. the underestimation at the low mass end at $z = 5.5$, can be explained by the fact that the CMFs given by the Parkinson et al. (2008) algorithm does not fully agree with the simulation as demonstrated in Figure 5.1. However, I find that this overestimation does not affect the stellar mass functions when applying a semi-analytic model to the augmented trees. I show this in Section 5.6.

The halo mass functions (HMFs) of the extended trees are demonstrated in the upper panels of Figure 5.4. They show excellent agreement with L35N2650. The lower panels of the figure explicitly show the HMFs of N-body, MC-I and MC-II halos from L105E10. As defined in Section 5.3.4, MC-I halos augment N-body merger trees, while MC-II halos are added to form a complete sample of halos, and are independent of N-body halos.

While MC-II halos dominate the population at lower redshifts, MC-I halos are the main contributor at higher redshifts. Hence, both types of halos are necessary to calculate the halo abundance across all redshifts.

5.4 Assigning halo positions

For modelling reionisation, I require spatial information for halos within the extended halo catalogues. I aim to assign a position to every Monte Carlo halo and ensure that their two-point statistics agree with N-body simulations. Section 5.4.1 discusses a random sampling method for placing MC-II halos within the simulation in the snapshot that the augmentation of the N-body merger trees is started, i.e. at $z = 5$. The method is also used to verify the results in Section 5.4.2, which introduces an approach for evolving the position of Monte Carlo halos based on the position of their descendant.

5.4.1 Populating halo positions

Monte Carlo halos can be populated into a simulation box using an analytic halo bias to transform the dark matter density field to a halo density field as a function of halo mass (de la Torre & Peacock, 2013; Angulo et al., 2014; Neyrinck et al., 2014; Ahn et al., 2015; Nasirudin et al., 2020). In this work, the dark matter density field is estimated from the L105N2048 N-body simulation using the nearest grid point method. The result is represented as a cubic grid. To estimate the halo density field, I adopt the non-linear halo bias proposed by Ahn et al. (2015), which avoids negative density in underdense regions, and results in better two-point correlation functions on smaller scales. Halo positions are obtained by random sampling. I normalise the halo density field derived from the halo bias, and treat it as a one-dimensional discrete probability distribution. Then, at a given snapshot, I assign every Monte Carlo halo to a cell according to this probability and place it uniformly within the cell so that the number of halos in each cell follows the Poisson distribution. This approach does not depend on the normalisation of the halo density field and can be applied to any given number of Monte Carlo halos.

To verify this method, I carry out a test within mass ranges that are well resolved by L105N2048. Specifically, I apply this method to 10^5 samples, placing them within an empty box and measuring their two-point correlation functions. Then, I compare the

results using N-body halos from L105N2048. I perform the test at $M_{\rm halo} = 10^{9.1}\,h^{-1}\,{\rm M}_{\odot}$ and $M_{\rm halo} = 10^{9.5}\,h^{-1}\,{\rm M}_{\odot}$ from $z = 5$ to $z = 10$ with different grid sizes. The results can be found in Figure 5.5, which shows good agreement with those estimated from the N-body simulation.

The small scale clustering predicted by the random sampling the method is affected by the choice of grid sizes. Halo positions within a cell of the grid are inaccurate since they are assumed to be uniformly distributed. As expected, the two-point correlations obtained using a 128^3 grid (shown as blue circles in Figure 5.5) are underestimated at separations smaller than $0.8\,h^{-1}$Mpc, which is equal to the cell size of the grid. In terms of the results using a 512^3 grid (grey circles), they have slightly larger clustering amplitudes over all scales than those using a 256^3 grid (red circles). A potential reason could be that the estimation of the dark matter density field becomes noisy when a larger number of cells are used. For the following applications, I adopt a 256^3 grid for the random sampling method. This choice is appropriate since the corresponding cell size ($0.4\,h^{-1}$Mpc) is smaller than the characteristic size of ionising regions (e.g. Furlanetto et al., 2006).

Unfortunately, I am unable to do the same test for Monte Carlo halos in the extended halo catalogues. This is because a complete sample of N-body halos at these mass ranges is only available in L35N2650, for which the box size is not sufficient to estimate two-point statistics. However, I note that the linearity of halo density fields increase towards lower halo mass, implying that the results are likely to be improved at $M_{\rm halo} \lesssim 10^{8}\,h^{-1}\,{\rm M}_{\odot}$. This argument indicates that the results in Figure 5.5 are conservative for estimating the accuracy of the method. Hence, my method can be safely applied to the mass ranges that are interested in this work.

5.4.2 Evolving halo positions

Evolution in the clustering of halos is influenced by their peculiar motions. My approach of evolving halo positions is based on the linear continuity equation. I again divide the L105N2048 box into a cubic grid with 256^3 cells. For Monte Carlo trees at t_1, the first step is to place the halos into the same cell as their direct descendant at t_2. I assume that the spatial distribution of the halos at t_1 can be described by a halo density field denoted as $\mathcal{D}(\vec{x}, t_1)$. The idea is to move these halos using a velocity field such that their

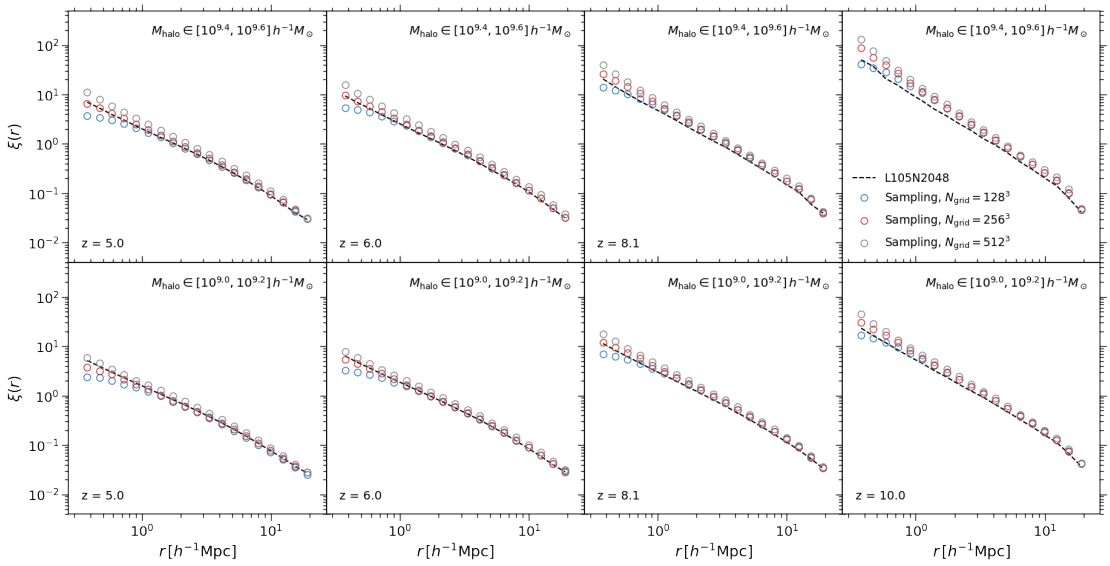


FIGURE 5.5: Comparison of two-point correlation functions produced using the random sampling method and estimated from N-body simulations. Empty circles are the results based on the random sampling method introduced in Section 5.4.1, with colours corresponding to different grid sizes as labelled on the top rightmost panel. Black dashed lines are estimated from the L105N2048 N-body simulations. Each row corresponds to a halo mass bin. These mass ranges are well resolved by L105N2048.

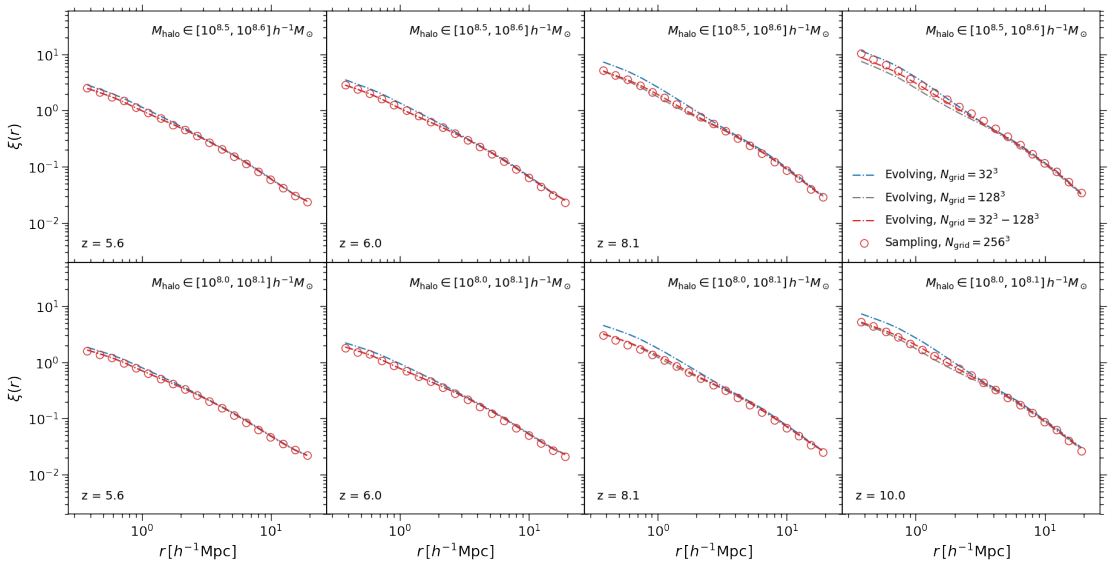


FIGURE 5.6: Comparison of two-point correlation functions produced using the evolving method and estimated from N-body simulations. Dash-dotted lines are the results based on the evolving method introduced in Section 5.4.2. For blue and grey lines, grids with 32^3 and 128^3 cells are used to calculate the velocity field respectively, while for red lines, the adopted grid size varies with redshift, with 128^3 cells at $z = 5 - 6$, 64^3 cells at $z = 6 - 8$, and 32^3 cells at $z > 8$. Red empty circles are the results obtained using the sampling method descibed in Section 5.4.1, which can be used to check the accuracy of the evolving method.

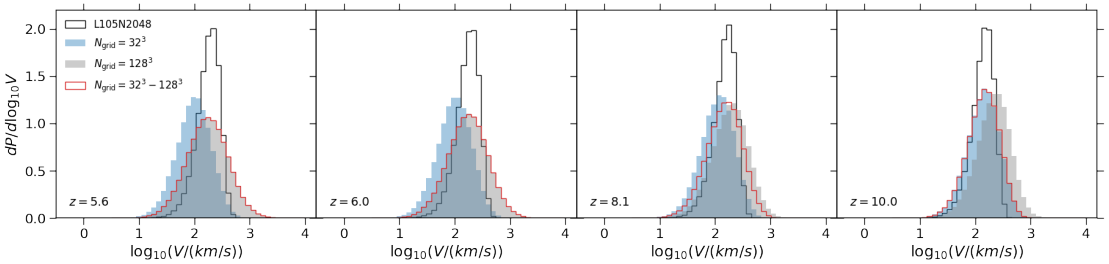


FIGURE 5.7: Peculiar velocity distributions of N-body and Monte Carlo halos. The velocities of Monte Carlo halos are derived using the method introduced in Section 5.4.2. For blue and grey histograms, grids with 32^3 and 128^3 cells are used in the calculations, while for red histograms, the adopted grid size varies with redshift, with 128^3 cells at $z = 5 - 6$, 64^3 cells at $z = 6 - 8$, and 32^3 cells at $z > 8$. The distributions of N-body halos are shown as black histograms.

spatial distribution becomes a desired halo density field denoted as $\mathcal{D}(\vec{x}, t_2)$. I assume that this process can be described by the linear continuity equation. If $\Delta t = t_2 - t_1$ is small, the velocity field can be obtained by

$$\nabla\vec{v}(\vec{x}, t_2) = -\frac{1}{\Delta t}\left[\mathcal{D}(\vec{x}, t_1) - \mathcal{D}(\vec{x}, t_2)\right] \tag{5.3}$$

In the linear regime, we want

$$\mathcal{D}(\vec{x}, t_1) = b(M_1, t_1)\delta_{\rm DM}(\vec{x}, t_1) \tag{5.4}$$

where M_1 is the mass of the Monte Carlo halos and $b(M, t)$ is the linear halo bias. After a forward evolution, the change of the density field for halos at t_1 with mass M_1 is contributed from both the variation of the background dark matter density field and local interactions such as smooth mass accretion and mergers. Although a detailed model that considers all the effects is complicated, I find that evolving halo positions using the following expression for $\mathcal{D}(\vec{x}, t_2)$ can lead to reasonable two-point statistics.

$$\mathcal{D}(\vec{x}, t_2) = b(M_1/\bar{\mu}_{\rm R}, t_2)\delta_{\rm DM}(\vec{x}, t_2), \tag{5.5}$$

where $\bar{\mu}_{\rm R}$ is the mean mass ratio.

Then, it is straightforward to compute the velocity field using the Fourier transform. The velocity field in k-space can be written as

$$\vec{v}(\vec{k}, t_2) = b(M_1/\mu_{\rm R}, t_2)\vec{u}(\vec{k}, t_2) - b(M_1, t_1)\vec{u}(\vec{k}, t_1) \tag{5.6}$$

with

$$\vec{u}(\vec{k},t) = \frac{i\vec{k}}{\Delta t k^2}\delta_{\rm DM}(\vec{k},t), \tag{5.7}$$

The real space velocity field then can be obtained using the inverse Fourier transform. Since $\vec{u}(\vec{k},t)$ is independent of halo mass, I only need to perform the Fourier transform once per snapshot, and the velocity can be calculated per halo, without any mass bins. This advantage is only available when the halo bias and the dark matter density field are separable. For the linear halo bias, I adopt the fitting model given by Tinker et al. (2010).

I apply this method to all extended halo catalogues and find that the choice of grid sizes to calculate $\vec{u}(\vec{k},t)$ can affect the results. In Figure 5.7, I show that the median velocity of Monte Carlo halos is underestimated at $z \sim 5$ using a 32^3 grid and is overestimated at $z \sim 10$ using a 128^3 grid. This trend is expected. The density field should not be over smoothed, which loses the information on density peaks. On the other hand, the halo bias increases rapidly with redshift, in which case the halo density field cannot be described by the linear bias. Smoothing the density field over larger regions contribute to increasing the linearity.

To verify the two-point correlation functions predicted by the evolving method, I have to use the sampling method introduced in the previous section. A direct comparison with L35N2650 is not feasible due to its limited box size, and the accuracy of this indirect approach is confirmed in the previous section. Figure 5.6 compares the two-point correlation functions obtained using the sampling and evolving methods. Since halo positions are evolved backwards with time, when a 32^3 grid is used, the errors due to the underestimation of the halo velocity cumulate towards higher redshifts, which results in the overestimation of the two-point correlation functions at $z \gtrsim 6$ (see blue dash-dotted lines). Overall, I find good agreement between the results based on both methods, particularly on large scales.

Based on the discussion above, I have decided to vary the grid size with redshift when evolving halo positions. Specifically, I use a 128^3 grid at $z = 5-6$, a 64^3 grid at $z = 6-8$, and a 32^3 grid at $z > 8$. This treatment results in both consistent two-point correlation functions and velocity distributions, which are shown as red dash-dotted lines and red histograms in Figure 5.6 and Figure 5.7 respectively.

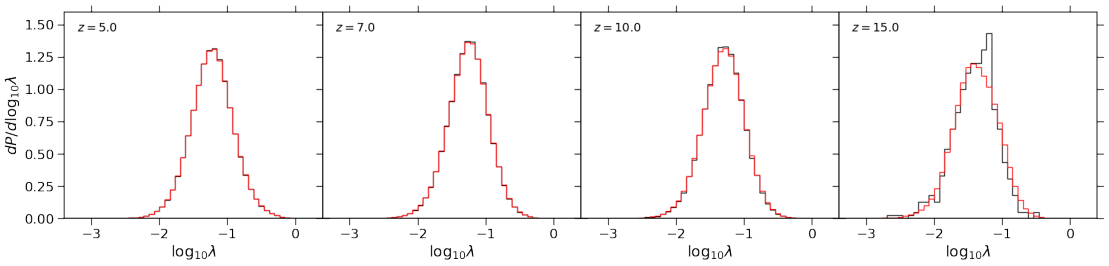


FIGURE 5.8: Spin distributions of N-body and Monte Carlo halos (back and red histograms respectively). The spin parameters of Monte Carlo halos are resampled from the distributions of N-body halos using a Gaussian kernel density estimator (see e.g. Scott, 2015). I only include N-body halos comprised of at least 100 particles to estimate their spin distributions, with subhalos excluded.

5.5 Spin parameters

Many semi-analytic models use the halo spin parameter (defined by Bullock et al. (2001)) to compute quantities including disk size and star formation rate. To facilitate this, I sample the spin parameter of Monte Carlo halos using the spin distributions estimated from the N-body simulation. At $z \geq 5$, negligible dependence on halo mass is found in the spin distributions of our simulations, which is consistent with Knebe & Power (2008) and Angel et al. (2016). The mass independent spin distributions can be described by a log-normal distribution (e.g. van den Bosch, 1998; Knebe & Power, 2008) or a modified profile taking into account the long tail of low spins (e.g. Bett et al., 2007; Angel et al., 2016). In this work, I adopt a non-parametric approach. I train a Gaussian kernel density estimator (see e.g. Scott, 2015) using samples from our N-body simulations (in $\log_{10}\lambda$ space), and assign the spin of Monte Carlo halos by resampling from the density estimator. I choose the bandwidth of the density estimator according to Scott's Rule (Scott, 2015).

In Figure 5.8, black and red histograms are the spin distributions based on N-body and Monte Carlo halos respectively. When assembling N-body halos to estimate the spin distributions, I only include halos comprised of at least 100 particles and exclude all subhalos. My results illustrate excellent agreement between the resampled and original distributions by construction. I note that my approach can be generalised to the case where spin parameter is tightly correlated with halo mass by splitting the total sample into several mass bins and applying the kernel density estimator to each subsample.

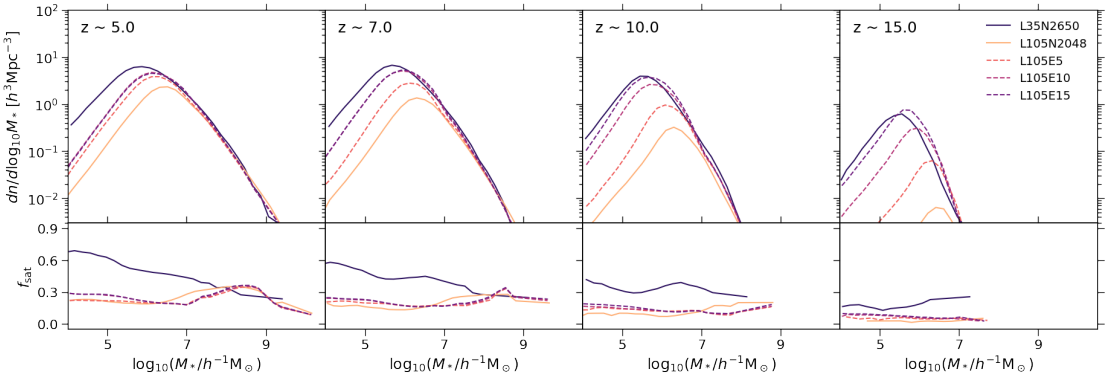


FIGURE 5.9: Stellar mass functions (upper panels) and satellite fractions (lower panels) predicted by the MERAXES semi-analytic model. For all panels, solid lines use the original halo merger trees from our N-body simulations. Dashed lines are the results based on extended catalogues, which consist of both N-body and Monte Carlo halos. Deeper colour corresponds to higher mass resolution. The information on each halo catalogue as labelled in the top right corner can be found in Table 5.3.

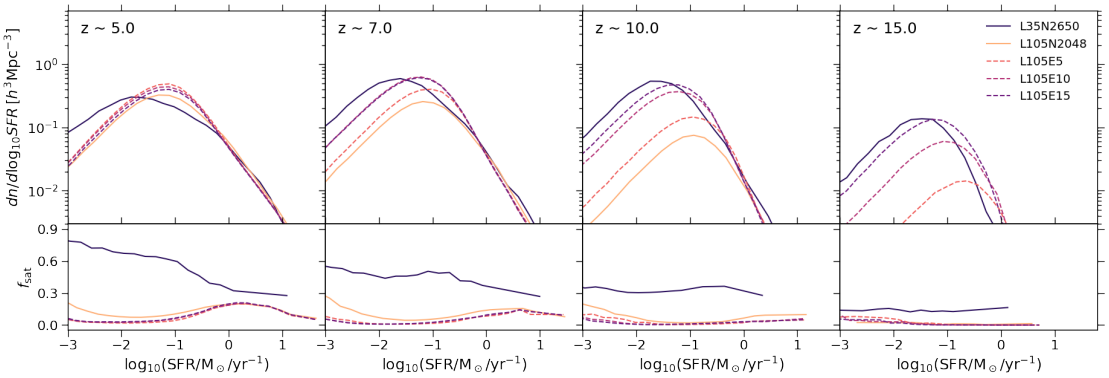


FIGURE 5.10: Star formation rate functions (upper panels) and satellite fractions as a function of star formation rate (lower panels) predicted by the MERAXES semi-analytic model. For all panels, solid lines use the original halo merger trees from our N-body simulations. Dashed lines are the results based on extended catalogues, which consist of both N-body and Monte Carlo halos. Deeper colour corresponds to higher mass resolution. The mass resolutions of L105E5, L105E10 and L105E15 are the atomic cooling thresholds at $z = 5$, 10 and 15 respectively. The information on each halo catalogue as labelled in the top right corner can be found in Table 5.3.

5.6 Application to Meraxes

I apply both the N-body and extended halo catalogues to the MERAXES semi-analytic model (Mutch et al., 2016). In addition to the implementation of several key galaxy formation processes including radiative cooling, star formation and supernova feedback, the MERAXES model is coupled with 21CMFAST (Mesinger & Furlanetto, 2007) to realise inhomogeneous reionisation feedback and to predict reionisation related properties such as the global neutral fraction and 21cm power spectra. The MERAXES model only seeds galaxies in halos whose mass is above the atomic cooling threshold. I adopt the same

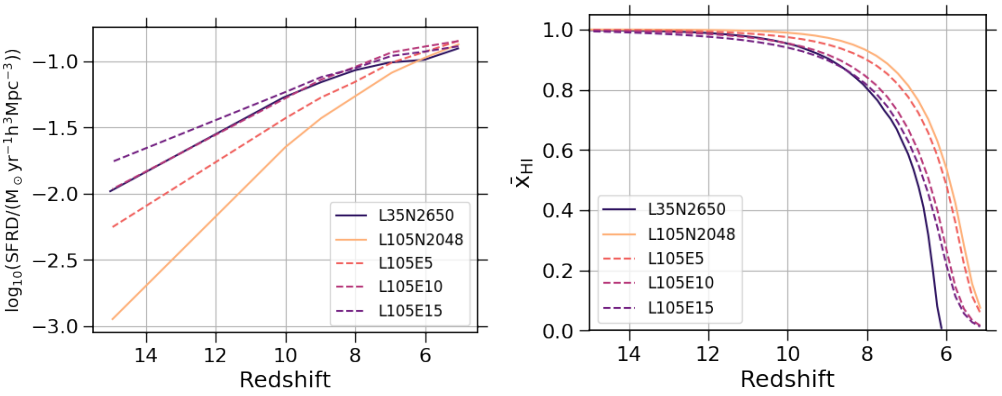


FIGURE 5.11: Star formation rate density (left panel) and volume-weighted neutral fractions (right panel) predicted by the MERAXES semi-analytic model. In all panels, solid lines use the original halo merger trees from our N-body simulations, and dashed lines are the results based on extended catalogues, which consist of both N-body and Monte Carlo halos. Deeper colour corresponds to higher mass resolution. The mass resolutions of L105E5, L105E10 and L105E15 are the atomic cooling thresholds at $z = 5$, 10 and 15 respectively. The information on each halo catalogue as labelled on the bottom can be found in Table 5.3.

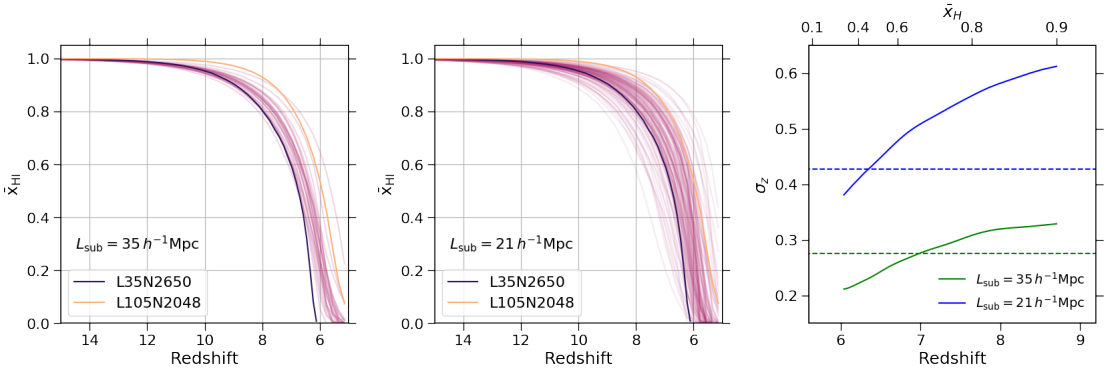


FIGURE 5.12: Effect of cosmic variance on the reionisation history. Left and middle panels show the reionisation histories in subvolumes with side lengths of $35\,h^{-1}$Mpc and $21\,h^{-1}$Mpc respectively. The latter is roughly equal to the maximum bubble size that is chosen for 21CMFAST. These results are based on L105E10. In right panels, solid lines show the standard deviations of redshift in subvolumes at fixed neutral fractions. Redshifts on the bottom axis are converted using the mean relation of the entire volume. The deviations are compared with the analytic predictions of Barkana & Loeb (2004), which are shown as dashed lines.

parameters as Mutch et al. (2016) but note that the model predictions can be different from Mutch et al. (2016) due to the use of different halo merger trees. However, the main focus of this work is to demonstrate the consistency between the N-body and extended halo catalogues and to illustrate the consequences of adopting different halo mass resolutions rather than to present a model that satisfies all current observational constraints.

5.6.1 Galaxy properties

Figures 5.9, 5.10 and 5.11 demonstrate the stellar mass functions, star formation rate functions, and star formation rate densities predicted by MERAXES respectively. L35N2650 is a small volume N-body simulation with very high mass resolution, which is used to verify the results based on the extended halo catalogues. Overall agreement can be found between the predicted galaxy properties using L35N2650 and extended trees up to $z \sim 10$, which are shown as purple solid and dashed lines respectively. I find a difference in the peaks of both the stellar mass and star formation rate functions, which may result from the fact that Monte Carlo merger trees do not contain subhalos [2]. This point is illustrated in the lower panels of Figures 5.9 and 5.10, where I show that L35N2650 provides significantly higher satellite fractions than the extended halo catalogues, particularly at the low stellar mass and low star formation rate ends. In MERAXES, all gas infalling into a friends-of-friends group is assumed to be accreted onto the central galaxy. Therefore, satellite galaxies have less fuel to form stars. Despite this disagreement, I find excellent agreement between the cosmic star formation rate densities obtained using L35N2650, L105E10 and L105E15 at $z < 10$. The result based on L105E15 shows higher star formation rate density than L35N2650 at $z > 10$. However, L35N2650 has a higher mass resolution. This is likely due to the overestimation of the halo mass functions at these redshifts as illustrated in Figure 5.4.

An additional finding is that the effect of mass resolution does not seem to be cumulative. While the mass resolutions of L105E5, L105E10 and L105E15 are different (and all above the atomic cooling threshold at $z = 5$), in Figure 5.9, their corresponding stellar mass functions overlap at $z = 5$. Figure 5.11 also shows that the star formation rate densities predicted by the extended trees converge towards $z = 5$. These indicate that if all halos above the atomic cooling threshold at a given redshift are resolved, an ability to resolve less massive halos at an earlier time has little effect on predicted galaxy properties such as the stellar mass and star formation rate functions at the given redshift.

[2] This may not be a problem for semi-analytic models that originally designed for Monte Carlo merger trees, which contain a treatment of subhalos. The MERAXES model used in this work is designed for N-body merger trees and does not implement the treatment.

5.6.2 Reionisation histories

Having demonstrated the consistency of galaxy properties using N-body and extended halo catalogues, I now focus on the predictions of cosmic reionisation. The end of reionisation is known as too rapid in simulations that do not resolve all faint galaxies or do not have a sufficiently large volume (Barkana & Loeb, 2004; Iliev et al., 2014). I therefore expect that the predictions of the reionisation history are sensitive to both halo mass resolution and simulation volume.

The right panel of Figure 5.11 illustrates the effect of halo mass resolution on the predicted volume-weighted neutral fractions. I see a difference between results using direct N-body merger trees (from L35N2650 and L105N2048). However, it is not straightforward to interpret this due to the different simulation volumes. My extended halo catalogues (L105E10 and L105E15) have the same volume as L105N2048 and produce consistent star formation rate densities with L35N2650. In the right panel, Figure 5.11 shows that the end of reionisation occurs earlier in L105E10 and L105E15 than in L105N2048, which confirms that mass resolution has an impact on the reionisation history. This is expected since the model assumed constant escape fraction for ionising photons. On the other hand, while having different mass resolutions, L105E10 and L105E15 predict similar reionisation histories. This implies that the convergence can be achieved by resolving the atomic cooling threshold at $z \sim 10$. Overall, my results suggest that resolving the atomic cooling limit at the beginning of reionisation is necessary to obtain convergent predictions for the reionisation history since reionisation is sensitive to cumulative star formation.

Small box simulations are known to suffer from both cosmic variance and lack of large scale modes (e.g. Barkana & Loeb, 2004). I demonstrate this effect using subvolumes of the L105E10 extended halo catalogue. In the left and middle panels of Figure 5.12, I show reionisation histories in two different sizes of subvolumes, having $L_{\rm sub} = 35\,h^{-1}$Mpc and $L_{\rm sub} = 21\,h^{-1}$Mpc. The former has the same volume as L35N2650, while the latter is roughly equal to the maximum bubble size that I choose in the 21CMFAST algorithm within MERAXES. Each subvolume contains different amounts of large scale power, leading to a rapid end of reionisation in each case, but at a range of redshifts. This explains the deviation of the shape of the late time reionisation history in L35N2650

from the predictions based on L105E10 and L105E15. The large volume simulations average cosmic variance shown within subvolumes in Figure 5.12.

In the right panel of Figure 5.12, I compare the standard deviation of redshift at fixed neutral fractions in the subvolumes (solid lines) with the analytic prediction of Barkana & Loeb (2004) (dashed lines). They pointed out that the difference of the collapse fraction in random regions of the Universe can be interpreted as an offset in redshift with respect to the cosmic mean. The scatter of the offset can be calculated from the critical collapse fraction, and be related to the width or duration of the reionisation history by equating it to the size of a particular reionisation region (Wyithe & Loeb, 2004). Despite the complexities in MERAXES, the analytic prediction provides a reasonable estimation of cosmic variance. Overall, my results reinforce the importance of a large volume for cosmic reionisation simulations, which has also been highlighted by previous studies (e.g. Iliev et al., 2006, 2014; Deep Kaur et al., 2020).

In addition, my results show that resolving all halos above the atomic cooling threshold across whole cosmic reionisation is important for calculating a converged reionisation history. Robertson et al. (2015) analysed the joint observational constraints of Thomson scattering optical depth measured by Planck Collaboration et al. (2016) and cosmic star formation rate density estimated by Madau & Dickinson (2014), suggesting that cosmic reionisation happens at $6 \lesssim z \lesssim 10$. My results imply that simulations should reach at least the atomic cooling threshold at $z = 10$ in order to explore such reionisation scenarios. The decrease of the atomic cooling threshold with increasing redshift places constraints on the required halo mass resolution of simulations towards the beginning of reionisation.

5.7 Summary

This work presents a hybrid method to compute high resolution halo merger trees within large volume N-body simulations for semi-analytic reionisation models, which is based on the work of Benson et al. (2016). As an application, I extend the mass resolution of halo merger trees extracted from the Genesis N-body $105\,h^{-1}$Mpc simulation box at $z \geq 5$. I verify the results using a small N-body simulation with very high resolution, and find good agreement for the halo mass functions. I also introduce a method to assign and

evolve the position of Monte Carlo halos. The resulting two-point correlation functions are consistent with N-body simulations at separations greater than $0.4\,h^{-1}$Mpc. The extended halo catalogues are then used as input for the MERAXES semi-analytic model with application to the reionisation history. My main findings can be summarised as follows:

- The predicted stellar mass functions, star formation rate densities and volume-weighted neutral fractions based on the extended halo catalogues are consistent with those in a high resolution compensated simulation.
- If all halos above the atomic cooling threshold at a given redshift are resolved, resolving even smaller halos at higher redshifts has negligible effect on predictions of galaxy population properties from the MERAXES semi-analytic model at the given redshift.
- The decreasing atomic cooling threshold requires simulations to have higher mass resolution towards higher redshifts. My model implies that the faint sources at the beginning of reionisation can have a significant impact on the reionisation history, and therefore reliable calculations of the reionisation history need the atomic cooling limit to be resolved throughout reionisation.
- The end of reionisation is predicted to be too rapid in simulations that either fail to resolve all faint galaxies or have a too small volume, putting demands on halo mass resolution and simulation volume. Using my extended tree algorithm, I show that the convergent predictions of the late stage reionisation history need both large volumes ($L_{\rm box} \gtrsim 100\,h^{-1}$Mpc) and resolution of the atomic cooling threshold across the whole reionisation history.

My methodology provides a powerful tool to achieve desired mass resolution in large volumes. The largest extended halo catalogue obtained in this work has the mass resolution at $M_{\rm halo} = 3.2 \times 10^7\,h^{-1}\,{\rm M}_\odot$ in a $105\,h^{-1}$Mpc box, equivalent to an N-body simulations with $\sim 6800^3$ particles. Given the efficiency of the Monte Carlo algorithms, my approach can be applied to larger volumes (several hundred Mpc on each side), which are necessary for studying the statistics of reionisation including X-ray heating and global 21cm signal during cosmic dawn.

Chapter 6

Summary

The epoch of reionisation is still an unknown era in the Universe. We have already observed many light sources from this epoch, including both galaxies and quasars. Future instruments, e.g. the James Webb Space Telescope (JWST) and the Square Kilometre Array (SKA), are expected to not only observe larger samples of light sources but also prob the ionising structure directly. The task for theoretical models is to interpret these observations. This thesis demonstrates that semi-analytic models provide a powerful tool to study both galaxy formation and cosmic reionisation physics. The three related topics explored in the thesis and their future improvements are summarised as follows:

1. The spatial distribution of galaxies can be described by the two-point correlation, which contains information on their host halo mass. Chapter 3 shows that the MERAXES semi-analytic model predicts consistent dependence of clustering strength on both UV-luminosity and stellar mass with observations. Such agreement validates the predicted correlation between halo mass and these galaxies properties. An improved approach is to treat the observed clustering as a constraint and use it to calibrate the model using the same method presented in Chapter 4. Such treatment was explored by van Daalen et al. (2016) in the local Universe.

2. Modelling dust extinction is a quite difficult task since intrinsic luminosity is required. Chapter 4 attempts to resolve the problem by determining intrinsic luminosity and dust extinction simultaneously with only UV observations. This approach takes full advantage of the fast evaluation speed of semi-analytic models and is very useful in the early Universes, where observations of infrared data are

very challenging. Future spectroscopic surveys from the JWST will be able to measure dust extinction using emission line ratios. Such results can be compared with those obtained in Chapter 4, which put further constraints on our model. An extension of the work is the incorporation of a dust evolution model (e.g Mancini et al., 2016; Popping et al., 2017a; Dayal & Ferrara, 2018), which allows the observations to put constraints on dust mass of galaxies.

3. A challenge in simulating cosmic reionisation is that both high mass resolution and large volume are required. Semi-analytic models have great potential to achieve this purpose, which, however, requires large N-body simulations as input. To overcome the challenge, Chapter 5 presents a hybrid method to produce large samples of halo merger trees. The effectiveness of this method is also demonstrated in the chapter. This work can be extended further by increasing the mass resolution further to the molecular cooling threshold, which is necessary for semi-analytic models that including Pop-III star formation (e.g. Visbal et al., 2018).

Bibliography

Ahn K., Iliev I. T., Shapiro P. R., Srisawat C., 2015, MNRAS, 450, 1486

Álvarez-Márquez J., et al., 2016, A&A, 587, A122

Anders P., Fritze-v. Alvensleben U., 2003, A&A, 401, 1063

Angel P. W., Poole G. B., Ludlow A. D., Duffy A. R., Geil P. M., Mutch S. J., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 459, 2106

Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2014, MNRAS, 442, 3256

Ashby M. L. N., et al., 2013, ApJ, 769, 80

Ashby M. L. N., et al., 2015, ApJS, 218, 33

Barisic I., et al., 2017, ApJ, 845, 41

Barkana R., Loeb A., 2004, ApJ, 609, 474

Barnes J., Hut P., 1986, Nature, 324, 446

Barone-Nugent R. L., et al., 2014, ApJ, 793, 17

Barry N., et al., 2019, ApJ, 884, 1

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Benson A. J., Cannella C., Cole S., 2016, Computational Astrophysics and Cosmology, 3, 3

Berger M. J., Colella P., 1989, Journal of Computational Physics, 82, 64

Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215

Bhatawdekar R., Conselice C., Margalef-Bentabol B., Duncan K., 2018, arXiv e-prints, p. arXiv:1807.07580

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bouchet F. R., Kandrup H. E., 1985, ApJ, 299, 1

Bouwens R. J., et al., 2011, ApJ, 737, 90

Bouwens R. J., et al., 2014, ApJ, 793, 115

Bouwens R. J., et al., 2015, ApJ, 803, 34

Bouwens R. J., et al., 2016, ApJ, 833, 72

Bower R. G., 1991, MNRAS, 248, 332

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240

Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, ApJ, 429, 582

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Capak P. L., et al., 2015, Nature, 522, 455

Ceverino D., Glover S. C. O., Klessen R. S., 2017, MNRAS, 470, 2791

Chabrier G., 2003, PASP, 115, 763

Charlot S., Fall S. M., 2000, ApJ, 539, 718

Collacchioni F., Cora S. A., Lagos C. D. P., Vega-Martínez C. A., 2018, MNRAS, 481, 954

Conroy C., 2013, ARA&A, 51, 393

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

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

Cousin M., Buat V., Lagache G., Bethermin M., 2019, arXiv e-prints, p. arXiv:1901.01747

Cowles M. K., Carlin B. P., 1996, Journal of the American Statistical Association, 91, 883

Cullen F., McLure R. J., Khochfar S., Dunlop J. S., Dalla Vecchia C., 2017, MNRAS, 470, 3006

Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265

Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1

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

Deep Kaur H., Gillet N., Mesinger A., 2020, arXiv e-prints, p. arXiv:2004.06709

Driver S. P., et al., 2018, MNRAS, 475, 2891

Duffy A. R., Wyithe J. S. B., Mutch S. J., Poole G. B., 2014, MNRAS, 443, 3435

Duncan K., et al., 2014, MNRAS, 444, 2960

Durkalec A., et al., 2018, A&A, 612, A42

Earl D. J., Deem M. W., 2005, Physical Chemistry Chemical Physics (Incorporating Faraday Transactions), 7, 3910

Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241

Elahi P. J., Poulton R., Canas R., 2019a, VELOCIraptor-STF: Six-dimensional Friends-of-Friends phase space halo finder (ascl:1911.020)

Elahi P. J., Poulton R., Tobar R., 2019b, TreeFrog: Construct halo merger trees and compare halo catalogs (ascl:1911.021)

Elahi P. J., Cañas R., Poulton R. J. J., Tobar R. J., Willis J. S., Lagos C. d. P., Power C., Robotham A. S. G., 2019c, PASA, 36, e021

Elahi P. J., Poulton R. J. J., Tobar R. J., Cañas R., Lagos C. d. P., Power C., Robotham A. S. G., 2019d, PASA, 36, e028

Fan X., et al., 2006, AJ, 132, 117

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778

Feroz F., Hobson M. P., 2008, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Finkelstein S. L., et al., 2012, ApJ, 756, 164

Finkelstein S. L., et al., 2015, ApJ, 810, 71

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Fudamoto Y., et al., 2017, MNRAS, 472, 483

Furlanetto S. R., McQuinn M., Hernquist L., 2006, MNRAS, 365, 115

Giavalisco M., 2002, ARA&A, 40, 579

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375

Greig B., Mesinger A., 2015, MNRAS, 449, 4246

Grogin N. A., et al., 2011, ApJS, 197, 35

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

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

Harikane Y., et al., 2016, ApJ, 821, 123

Harikane Y., et al., 2018, PASJ, 70, S11

Hassan S., Davé R., Finlator K., Santos M. G., 2016, MNRAS, 457, 1550

Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, MNRAS, 396, 535

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

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663

Hildebrandt H., Pielorz J., Erben T., van Waerbeke L., Simon P., Capak P., 2009, A&A, 498, 725

Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, MNRAS, 369, 1625

Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725

Illingworth G. D., et al., 2013, ApJS, 209, 6

Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, MNRAS, 442, 1805

Ishikawa S., Kashikawa N., Toshikawa J., Tanaka M., Hamana T., Niino Y., Ichikawa K., Uchiyama H., 2017, ApJ, 841, 8

Johnson J. L., Dalla Vecchia C., Khochfar S., 2013, MNRAS, 428, 1857

Jose C., Lacey C. G., Baugh C. M., 2016, MNRAS, 463, 270

Kampakoglou M., Trotta R., Silk J., 2008, MNRAS, 384, 1414

Katz H., et al., 2020, MNRAS,

Kauffmann G., 1996, MNRAS, 281, 475

Knebe A., Power C., 2008, ApJ, 678, 621

Koekemoer A. M., et al., 2011, ApJS, 197, 36

Koprowski M. P., et al., 2018, MNRAS, 479, 4355

Kroupa P., 2002, Science, 295, 82

Labbé I., Bouwens R., Illingworth G. D., Franx M., 2006, ApJ, 649, L67

Labbé I., et al., 2015, ApJS, 221, 23

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011, MNRAS, 416, 1566

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573

Landau L. D., Lifshitz E. M., 1975, The classical theory of fields

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lee K.-S., Giavalisco M., Gnedin O. Y., Somerville R. S., Ferguson H. C., Dickinson M., Ouchi M., 2006, ApJ, 642, 63

Leitherer C., et al., 1999, ApJS, 123, 3

Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, ApJS, 189, 309

Leitherer C., Ekström S., Meynet G., Schaerer D., Agienko K. B., Levesque E. M., 2014, ApJS, 212, 14

Ling E. N., Barrow J. D., Frenk C. S., 1986, MNRAS, 223, 21P

Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 235

Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113

Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140

Ma X., et al., 2019, MNRAS, 487, 1844

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Maio U., Ciardi B., Dolag K., Tornatore L., Khochfar S., 2010, MNRAS, 407, 1003

Mancini M., Schneider R., Graziani L., Valiante R., Dayal P., Maio U., Ciardi B., 2016, MNRAS, 462, 3130

Mason C. A., Trenti M., Treu T., 2015, ApJ, 813, 21

Merlin E., et al., 2016, A&A, 595, A97

Mesinger A., Furlanetto S., 2007, ApJ, 669, 663

Meurer G. R., Heckman T. M., Calzetti D., 1999, ApJ, 521, 64

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

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

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJ, 638, L51

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691

Mutch S. J., Poole G. B., Croton D. J., 2013, MNRAS, 428, 2001

Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250

Narayanan D., Davé R., Johnson B. D., Thompson R., Conroy C., Geach J., 2018, MNRAS, 474, 1718

Nasirudin A., Iliev I. T., Ahn K., 2020, MNRAS, 494, 3294

Neal R. M., 1996, Statistics and Computing, 6, 353

Neistein E., Dekel A., 2008, MNRAS, 383, 615

Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, MNRAS, 441, 646

Ocvirk P., et al., 2016, MNRAS, 463, 1462

Oesch P. A., et al., 2010, ApJ, 714, L47

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Ono Y., et al., 2018, Publications of the Astronomical Society of Japan, 70, S10

Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544

Park J., Kim H.-S., Wyithe J. S. B., Lacey C. G., Baugh C. M., Barone-Nugent R. L., Trenti M., Bouwens R. J., 2016, MNRAS, 461, 176

Park J., et al., 2017, MNRAS, 472, 1995

Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933

Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

Peebles P. J. E., 1980, The large-scale structure of the universe

Pillepich A., et al., 2018, MNRAS, 475, 648

Planck Collaboration et al., 2016, A&A, 594, A13

Poole G. B., Angel P. W., Mutch S. J., Power C., Duffy A. R., Geil P. M., Mesinger A., Wyithe S. B., 2016, MNRAS, 459, 3025

Poole G. B., Mutch S. J., Croton D. J., Wyithe S., 2017, MNRAS, 472, 3659

Popping G., Somerville R. S., Galametz M., 2017a, MNRAS, 471, 3152

Popping G., Puglisi A., Norman C. A., 2017b, MNRAS, 472, 2315

Portinari L., Chiosi C., Bressan A., 1998, A&A, 334, 505

Press W. H., Schechter P., 1974, ApJ, 187, 425

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Qin Y., et al., 2017, MNRAS, 472, 2009

Qin J., Zheng X. Z., Wuyts S., Pan Z., Ren J., 2019a, MNRAS, 485, 5733

Qin Y., Duffy A. R., Mutch S. J., Poole G. B., Mesinger A., Wyithe J. S. B., 2019b, MNRAS, 487, 1946

Ritter C., Côté B., Herwig F., Navarro J. F., Fryer C. L., 2018, The Astrophysical Journal Supplement Series, 237, 42

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19

Roche N., Eales S. A., 1999, MNRAS, 307, 703

Rogers A. B., et al., 2014, MNRAS, 440, 3714

Rosdahl J., et al., 2018, MNRAS, 479, 994

Safarzadeh M., Hayward C. C., Ferguson H. C., 2017, ApJ, 840, 15

Saitoh T. R., 2017, AJ, 153, 85

Salpeter E. E., 1955, ApJ, 121, 161

Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, MNRAS, 431, 1366

Schaye J., et al., 2010, MNRAS, 402, 1536

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

Scott D. W., 2015, Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons

Shaw J. R., Bridges M., Hobson M. P., 2007, MNRAS, 378, 1365

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Shi Y., Eberhart R., 1998, in 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360). pp 69–73

Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, American Institute of Physics Conference Series Vol. 735, American Institute of Physics Conference Series. pp 395–405, doi:10.1063/1.1835238

Sobacchi E., Mesinger A., 2013, MNRAS, 432, L51

Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, MNRAS, 423, 1992

Song M., et al., 2016, ApJ, 825, 5

Speagle J. S., 2019, arXiv e-prints, p. arXiv:1904.02180

Springel V., 2005, MNRAS, 364, 1105

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

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878

Trac H. Y., Gnedin N. Y., 2011, Advanced Science Letters, 4, 228

Vázquez G. A., Leitherer C., 2005, ApJ, 621, 695

Visbal E., Haiman Z., Bryan G. L., 2018, MNRAS, 475, 5246

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230

Weinberg S., 1973, American Journal of Physics, 41, 598

Windhorst R. A., et al., 2011, ApJS, 193, 27

Wise J. H., Turk M. J., Norman M. L., Abel T., 2012, ApJ, 745, 50

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560

Wyithe J. S. B., Loeb A., 2003, ApJ, 595, 614

Wyithe J. S. B., Loeb A., 2004, Nature, 432, 194

Wyithe J. S. B., Loeb A., 2013, MNRAS, 428, 2741

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019, MNRAS, 483, 2983

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., Venkatesan A., Behroozi P., Ferguson H. C., 2020, MNRAS,

Zhang J., Fakhouri O., Ma C.-P., 2008, MNRAS, 389, 1521

da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595

da Cunha E., Eminian C., Charlot S., Blaizot J., 2010, MNRAS, 403, 1894

de la Torre S., Peacock J. A., 2013, MNRAS, 435, 743

van Daalen M. P., Henriques B. M. B., Angulo R. E., White S. D. M., 2016, MNRAS, 458, 934

van den Bosch F. C., 1998, ApJ, 507, 601

van der Burg R. F. J., Hildebrandt H., Erben T., 2010, A&A, 523, A74