UNIVERSITY OF SOUTHAMPTON

Growing self consistent galaxies in empirically modelled environments using STEEL: the STatistical sEmi-Empirical modeL

by

Philip John Grylls

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

in the

Faculty of Engineering and Physical Sciences

Physics and Astronomy

June 2020

UNIVERSITY OF SOUTHAMPTON

Abstract

Faculty of Engineering and Physical Sciences
Physics and Astronomy

Doctor of Philosophy

by Philip John Grylls

Several popular techniques exist for cosmological modelling of galaxies, from computationally intensive hydrodynamical simulations, ab-initio semi-analytic models, to data-driven semi-empirical models. Each has caveats that limit the range of predictions they can make. In hydrodynamical (sub-grid) and semi-analytic models, the high parametrisation can cause degeneracies that prevent clear analysis of which processes are driving galaxy evolution. Furthermore, all must balance simulating a large universe against simulating galaxies with high fidelity. This trade-off between volume and resolution introduces a bias in the number of objects simulated at different masses given differences in cosmological abundances.

This thesis describes the STatistical sEmi-Empirical model, STEEL, and its contributions to the modelling of galaxies. STEEL is built to overcome the limitations of volume and resolution. We remove these constraints and biases using a novel technique to replace discrete haloes with a statistical alternative. This statistical dark matter backbone is then combined with empirical techniques, e.g. Abundance Matching, to create STEEL.

In this thesis, we use STEEL to empirically generate robust assembly histories of galaxies, constrained using SDSS and high redshift cluster observations. Using these constraints we probe the in-situ vs. ex-situ growth of galaxies. It is found that within ΛCDM hierarchical assembly using certain stellar mass functions to populate haloes produces a satellite accretion history that is inconsistent with the central galaxy growth.

Furthermore, there is a noted tension in the observed galaxy pair fraction and its evolution with redshift. We use the flexible nature of STEEL we present a systematic investigation into how stellar mass function derivations affect the pair fraction. It is found that the pair fraction can be substantially altered by the type of stellar mass estimates, providing an avenue to remove the discrepancies in pair fraction observations.

Finally, there is still an active debate over the significance of mergers on the morphological evolution of galaxies. We find that mergers are capable of creating the observed elliptical fractions and, additionally, a two-pathway merger and in-situ disk instability can produce the observed lenticular fractions.

In summary, the STatistical sEmi-Empirical modeL, STEEL is a new take on galaxy modelling. In this thesis STEEL has been used to add constraints to galaxy assembly histories, satellite distributions, star formation rates, and pair fractions. Working alongside new extra-galactic surveys, such as EUCLID, STEEL has the potential to be a prominent feature in the future of extra-galactic astrophysics.

The reason cosmological modellers should never work *too* hard:
"There is a theory that states that if ever anyone discovers exactly what the Universe is for it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory that states this has already happened."

Douglas Adams, The restaurant at the end of the Universe.

Contents

List of Figures

List of Tables

Abbreviations

(S)HMF	(**Sub**) **H**alo **M**ass **F**unction
ΛCDM	Λ **C**old **D**ark **M**atter
(s)SFR	(**s**pecific) **S**tar **F**ormation **R**ate
SMF	**S**tellar **M**ass **F**unction
SMHM	**S**tellar-**M**ass-**H**alo-**M**ass
STEEL	**ST**atistical s**E**mi-**E**mpirical mode**L**
PJG	**P**hilip **J**ohn **G**rylls
FS	**F**rancesco **S**hankar
SP	**S**ai **P**andian

Declaration of Authorship

I, Philip John Grylls, declare that this thesis titled, 'Growing self consistent galaxies in empirically modelled environments using STEEL: the STatistical sEmi-Empirical modeL' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 28/04/2020

Acknowledgements

- Charlotte Louise Grylls, my loving and supportive wife, without whom none of this would be possible.
- Jamie Francis Friel, my original co-author and best friend from day one. Your enthusiasm and friendly competition in science and mathematics have helped me endlessly.
- Helen Frances and John Grylls, my parents who encouraged my science from my first breath and endured the tedious conversations that have yet to stop. I owe them both more than I can ever repay.
- Molley Scoble, Liam Jones, Rebecca Pembroke, and Kieran Neville, who have kept me sane throughout my education. Also to all my friends not mentioned thank you for being there to share a pint and ignore me talking endlessly about science.
- My schools St. Louis Catholic School Frome and Prior Park College Bath and the excellent teachers who fostered environments that created a love of learning and an appreciation of academic excellence. Specifically Deirdre Cromie, a dedicated teacher who laid much of the foundation of my formal education and whom I distinctly remember arguing at length about triangles with, and Sarah Smith whom I regard as one of the greatest scientific influences in my life.
- My friends and colleagues in the Southampton Astronomy Department, even if only Lorenzo Zanizi studies the correct objects!
- Finally, Dr Francesco Shankar, his passion for astronomy and this work has been outstanding throughout.

Dedicated to my wife, my friends, and my family.

Chapter 1

Introduction

Say first, of God above, or man below,
What can we reason, but from what we know?
Of man what see we, but his station here,
From which to reason, or to which refer?
Through worlds unnumber'd though the God be known,
'Tis ours to trace him only in our own.
He, who through vast immensity can pierce,
See worlds on worlds compose one universe,
Observe how system into system runs,
What other planets circle other suns,
What varied being peoples ev'ry star,
May tell why Heav'n has made us as we are.
But of this frame the bearings, and the ties,
The strong connections, nice dependencies,
Gradations just, has thy pervading soul
Look'd through? or can a part contain the whole?

Is the great chain, that draws all to agree,
And drawn supports, upheld by God, or thee?
- Alexander Pope 1733, 'An Essay on Man: Epistle I'

1.1 Motivation

1.1.1 The answer we seek.

The firmament has been the muse of humans for as long as we have recorded our history and most likely longer still. The field of astronomy is descended from priests who worshipped celestial objects as the divine and sought for them to bring meaning to their

world. Structures such as Stonehenge use celestial objects to allow people to track the 'repetitious' passage of time, thus being able to predict the seasons. It is possible that this correlation between the heavens and the earthly systems so critical to life is the reason that humanity became convinced that the universe was there for our benefit. Elaborate orrerys with the earth at the centre of all creation were built to explain how the sun and planets orbit around us cementing our belief that we are at its origin. The progression to modern astronomy was slow. However, thanks to advances in scientific thinking and technological process we have shed many of the biases and limitations that once held us back. We are now able to discover our place in cosmos that is unfathomably enormous, diverse and inhospitable.

The Milky Way is our home and was the first observed galaxy. The name comes from Greek to mean 'milky circle' stemming from our belief that the universe exists to give us life and gives the galaxy a mammalian nurturing characteristic. The idea of complexity emerging from the universe is present in the first classification of the structures of galaxies by Edwin Hubble [24, 25]. Complex spirals were thought to be the descendants of elliptical galaxies leading to the misnomer of 'early-type' (elliptical) and 'late-type' (spiral) galaxies, as it has since been found that, amongst other formation pathways, it is spiral galaxies that transform into elliptical galaxies. The Universe is at its core in antithesis to the way humanity regards themselves, it thus presents a challenge in thinking to detach hubris and think logically about a system that in its entirety is truly incomprehensible. This must be reflected in galactic modelling we must understand the limitations of our work and the scope of what each model can explain.

Most advances in astronomy have come from technological progress. When the most advanced way to view the cosmos was with the naked eye we observed little and as such interpretations were limited. With the advent of optics, such as lenses and then mirrors, which led to the building of telescopes, the ability to look deeper into space became a possibility and the first work on the classification of astrophysical objects was carried out. Modern advances such as CCD/SED photographic plates, and telescopes that work outside of the visible wavelengths, gave information far beyond what the human eye can see.

With the influx of new information, galaxy models have become further refined. The emergent complexity proposed by Hubble is superseded by a hierarchical formation which mutates structures. Observations of our galaxy and distant galaxies show how different morphologies of galaxies are more or less common at different masses. Galaxies with different masses and morphologies can be seen to form stars at different rates, with some of the most massive appearing 'red and dead'. We also find an 'arrow of time' where galaxies seem to increase in star formation peak and drop showing we are likely past the point of maximum stellar mass creation in the Universe. Whilst these discoveries are far

more complex then those made by studying the sky only with the naked eye, we remain connected to our predecessors by the ultimate pervasive question of 'why?'[1].

1.1.2 The tools we use.

The field of astronomy is both privileged and restricted by observations. As no experiments can be performed to generate new data the traditional theory, experiment, data, analysis, theory, ... cycle that exists in most other fields is not possible. It is such that theoretical models are often driven by the data and must be careful not to over-fit. Furthermore, for astronomers who observe galactic scale objects, the timescales involved eclipse not only a human lifespan but the entirety of human civilisation. It is then fortunate that the finite speed of light in conjunction with the massive distances between galaxies mean observations of the most distant galaxies are also observations of the history of galaxies. Therefore, by observing galaxies at different epochs we can construct a picture of how the galactic population has changed and evolved. Under the standard axiom that the Universe is, at large scales, homogeneous and isotropic, we can assume that the galaxies observed in the far distance, in both time and space, are similar to the progenitors of observed local galaxies that are closer in time and space. Using observations of the start-point and end-point of galaxy populations theories of how galaxies evolve can be constrained. To test these theories the ability to model galaxies and 'fast forward' time is required such that simulations can provide results in a reasonable time frame.

The first such model by Holmberg [26] predates digital computing and used an array of light bulbs where the 'candle power', or flux, was a proxy for mass. Two arrays of bulbs were constructed and by measuring the intensity of light along two axes the total gravitational attraction was calculated. The light is a reasonable proxy for mass due to the similarity in the inverse square laws that govern both the spread of light and an increase in the gravitational potential. Each mass element (light bulb) is then given an acceleration and its position updated manually. In the Holmberg [26] model the merging of two systems of distributed mass is confirmed along with an investigation of the change of shape of the systems.

With the rapid increase in computing power over the last century, the capability of galaxy simulations has grown. The simulation of mass and mergers have improved with more and smaller 'mass elements'. Beyond simply testing the dynamics of mass, simulations also have prescriptions for the formation of stars from gas, the feedback of energy from central black holes millions of times the mass of our sun, feedback from supernovae, and more... [e.g., 27, 28]. Despite these major advancements made over many decades, a full understanding of the assembly and evolution of galaxies in our Universe is beyond our

[1] 42

capability. Other less computationally intensive analytic tools have been used to model from 'first principles', in this instance gas collapse, the formation of galaxy populations. Since their original inception where galaxies form from gas via the loss of angular momentum and cooling to form galactic disks [29], these so-called "Semi-Analytic" Models have branched out to cover tens of different analytic recipes that try to balance different processes to faithfully recreate the diversity in galaxy populations [e.g., 30, 31]. A recent development in galactic modelling attempts not to recreate the entire Universe from first principles but to use what we observe as a guide providing powerful constraints. These "Semi-Empirical Models" are observation-driven and ask focused questions to understand if a given hypothesis/model of galaxy formation is adequate to 'link' the observed galactic population over cosmic history [8, 17, 32–34].

1.1.3 The questions we answer.

In this thesis, we describe STEEL , the STastical sEmi-Empirical modeL, PJG's contribution to the galactic modelling community. STEEL has been designed to be complementary to existing models of and approaches to galaxy formation. STEEL uses semi-empirical modelling techniques but its defining characteristic is its statistical nature. Traditional models, described in more detail below, simulate discrete elements of dark matter extracted from a cosmological volume. STEEL instead creates a simulation of galaxy populations using number density functions to describe cosmological volumes. By design, this avoids the volume and resolution constraints that effects discrete object simulations. The full methodology of STEEL is given in Chapter 2. The advantage of having a volume free model is twofold. Firstly, we can simulate rare objects that have a number density far lower than what can be extracted from traditional models. Secondly, STEEL is not biased in favour of smaller objects which have higher number densities and are thus simulated orders of magnitude more often in traditional simulations.

1.2 ΛCDM Cosmology

Dark Matter is a theorised form of matter that interacts only through gravity and is therefore not observable with traditional methods that rely on photons. As dark matter is five times more abundant than normal matter it is the dominant gravitational force in the universe. This has important implications for the expected galaxy populations as in a ΛCDM universe baryonic matter is theorised to trace the dark matter distribution. Galaxies are thought to assemble at the centre of dark matter haloes where gas accumulates, the larger the halo the larger quantity of gas collects in the halo and the larger the galaxy. Hierarchical assembly of dark matter haloes is then translated into the galaxy population. Galaxies will follow the merger history of their host haloes, smaller galaxies will follow their host halo when it merges with a larger halo. Massive galaxies,

which reside in massive haloes that have accreted many other haloes through mergers, are expected to be surrounded by many smaller satellite galaxies.

The first notion of dark matter was from Zwicky [35], who observed that the binding mass required to hold the Coma Cluster together was roughly 400 times the observable total stellar mass. It took many years and further evidence such as observations of the motions of satellites around our galaxy before the ideas of dark matter became mainstream. Furthermore, observations of the rotation curves of galaxies [36] were found to support a 'dark' component; the rotation of the outer regions of the galaxy did not fall off as the mass profile of the luminous galaxy would suggest within a Newtonian framework of gravity. Alternative theories to explain the discrepancies from Newtonian gravity exist, such as Modified Newtonian Dynamics (MOND) that suggests gravity acts differently at large scales. However, dark matter is now the preferred model amongst physicists to explain the aforementioned observations.

As the observations of mass deficit can be made on all galaxies it follows that dark matter must permeate the entire universe. Additionally, it interacts only through gravity giving no electromagnetic signature. It is still unclear what dark matter 'is' but through simulations, observations, and experiments we can put constraints on its possible nature. Initial theories suggested brown dwarf stars or black holes, dark yet familiar objects that satisfy both requirements of adding mass and being mostly invisible, however these were quickly dismissed [37]. It is now thought that dark matter consists of a massive subatomic particle characterised by low thermal velocity, and is hence known as cold dark matter (CDM). Including the cosmological constant (Λ) required for a flat universe [38, 39], alongside this model for dark matter, the leading cosmological theory of ΛCDM was developed.

The prediction from ΛCDM of foremost importance to galactic modelling is that of *hierarchical assembly*. Dark matter collapses under gravity from an initial density field, that mirrors density fluctuations in the cosmic microwave background. This means that areas of greater density collapse faster and eventually attract other collapsed regions further increasing their mass. This structure has become commonly referred to as the 'cosmic web'. As dark matter interacts only via gravity it is relatively easy to simulate on large scales. There have been many massive ΛCDM simulations using various cosmologies (WMAP5, Planck), of note are the Millennium simulation [40] and the Bolshoi/MultiDark simulations [41]. We show a visualisation of the Bolshoi Plank dark matter simulation in Figure 1.1. The 'cosmic web' is composed of matter haloes connected by filamentary structures seen here as the darkest regions. There is a notable void in the upper right of the image and two clusters, groups of many dark matter haloes in one spatial location, one in the top middle and one in the bottom left. The initial dark matter distribution is accentuated by giga-years of gravitational collapse creating the complex structure seen in Figure 1.1.

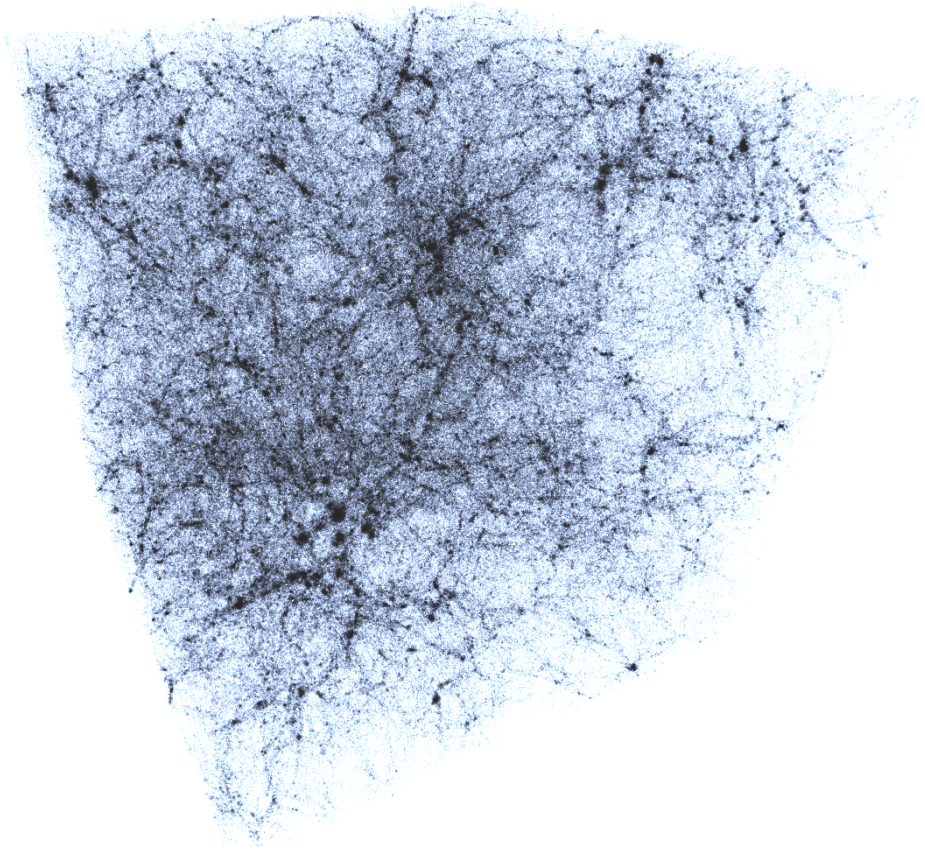


FIGURE 1.1: A visualization of the 'cosmic web'. The shading shows the distribution of dark matter. The darkest regions are the spherical 'haloes'. Between the haloes extended filaments can be seen. Regions of note are the void in the upper right and the clusters in the top middle and lower left. Initially the dark matter density field would have been similar to the microwave background but giga-years of collapse have accentuated small features into a large structure. Image Credit: C.Marsden using the galaxy and structure visualization code Astera processing the Bolshoi Planck dark matter simulation (https://astera.soton.ac.uk/).

In order to analyse complex dark matter simulations, several classifications have been developed to describe the dark matter haloes. Firstly, 'virialised haloes' collections of dark matter particles[2] which are gravitationally interacting and are balanced in terms of their potential and kinetic energy such that the group is neither expanding nor contracting. Secondly, from a structural point of view, haloes can be defined by their overdensity. Specifically, a halo is defined as a sphere containing a density of mass that is a set number times that of a reference average density. The average density definitions commonly used are defined by either the background[3] or critical[4] density. For example M_{200c} would be the mass of a halo defined as the spherical region that is 200 times the critical density. To identify these regions a number of methods have been developed:

[2]Note here a dark matter particle is a simulation element, often millions of solar masses, and not the theorised fundamental particle.

[3]The average density of the Universe.

[4]The density required such that the universe will not expand forever.

- Bound Density Maximum: BDM classifies halo structure by defining a spherical overdensity threshold, e.g. M_{200c}, then removes particles that exceed the escape velocity of the halo mass [42].

- Friends of Friends: FOF uses a dark matter particle linking threshold, particles are 'linked' together if they are closer than the threshold. One particle cannot be in two FOF groups simultaneously such that particle groups are unique [43].

- ROCKSTAR (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement): ROCKSTAR is a cutting-edge development of the FOF halo finders. Using a 6-dimensional phase space and a temporal dimension, the structure and evolution of haloes are extracted from dark matter simulations [44].

- MHF (MLAPM's halo finder): MHF uses the adaptive mesh of MLAPM to find haloes and substructure. The adaptive mesh naturally traces the density profile of the halo structure with the gurd structure surrounding areas of high density. This method avoids the scale lengths found in FOF [45].

Once equipped with a halo classification and an extraction tool, a language and formalism to describe how these structures relate to one another is required. Hierarchical assembly predicts halo growth via accretion followed by cannibalisation of a smaller halo. This process is not instantaneous and during this time the smaller halo is situated within the larger and can be referred to as a subhalo. Subhaloes then may contain further structure from smaller haloes they have accreted. Thus there are different 'orders' of haloes: 0^{th} order or central halos, 1^{st} order or subhaloes whose parent is a 0^{th} order halo, 2^{nd} order or subhaloes whose parent is a 1^{st} order halo, and so on. Useful statistical descriptions of the haloes are the halo and subhalo mass functions. The halo mass function is defined as the number density of central haloes of a given mass per unit volume. The subhalo mass function has multiple definitions depending on the definition of subhalo, each definition is useful to a different context. Firstly, one can define the global subhalo mass function, similarly to the halo mass function this is the number density of subhaloes at a given mass per unit volume. Secondly, one can define the subhalo mass function relative to a particular central host halo, which provides the number of subhaloes of a given mass ratio $M_{subhalo}/M_{central}$ one expects to be associated to a halo mass $M_{central}$.
Once a subhalo has been accreted onto a central halo it begins the process of merging. As the subhalo orbits, it gradually loses mass to the parent halo due to tidal disruption. This process is called 'stripping', initially the outer regions of the subhalo are stripped with the denser centre surviving for longer periods. Once a subhalo can no longer be resolved by the simulation, i.e. its mass becomes indistinguishable from the parent halo's mass, it is regarded as fully merged. Information on this process is derived from using the aforementioned halo finders to track substructures between simulation snapshots.

As this is a simulated result it is important to note that the simulation calibration, e.g. the number of particles and force softening, can strongly affect the timescale of subhalo disruption [46].
As the subhaloes are an evolving population, it is necessary to state which subhalo mass definition one is using:

- Unevolved Subhalo Mass Function (USHMF): This mass function is the total number of subhaloes that have accreted onto a central halo over its entire history. The subhalo mass is defined at the point of accretion and is frozen at infall and no number density is ever lost. This mass function is useful for understanding the total accretion history of a halo.

- Unevolved Surviving Subhalo Mass Function (USSHMF): This mass function is the total number of subhaloes that reside in a halo at a given epoch. Subhaloes once accreted are frozen in mass, however, at the time when they would have fully merged are then considered removed from the count. This mass function is useful as it retains information about the infall masses and can be further categorised to retain information about number density contribution from different infall epochs.

- Evolved Subhalo Mass function (ESHMF): This mass function is the total number of subhaloes that reside in a halo at a given epoch where the subhaloes have undergone mass loss/transfer to the central halo. This mass function is the 'true' subhalo mass function that one would expect to see if dark matter were directly testable, e.g. via galaxy lensing in groups and clusters [47].

In Figure 1.2, we show the halo mass function at redshift $z = 0.1$, black line. Note the characteristic Schechter function shape with an inverse linear relationship between number density and halo mass in log-log space, which breaks and rapidly declines above a threshold mass, which at this redshift is roughly $M_h \sim 10^{14}\ M_\odot$. We also show the Global Unevolved Surviving Subhalo Mass Function at redshift $z = 0.1$ the coloured shading represents the contribution to the global USSHMF at different redshifts of infall, note the logarithmic scaling means the shading is not linear. Recent accretion is the dominant contributor of subhaloes at all masses. Furthermore, high mass subhaloes $M_h \geq 10^{13} M_\odot$ accrete only at redshifts below $z < 2$.
In addition to the distinction between central and satellite haloes, information about the full mass assembly is also extracted from simulations. Halo assembly histories are commonly visualised as simple(*ish*) tree networks, referred to as a "merger trees". Central haloes are identified at a low redshift and the main progenitors followed backwards in time to create the central 'trunk'. At any epoch where a merger occurs, the main progenitor

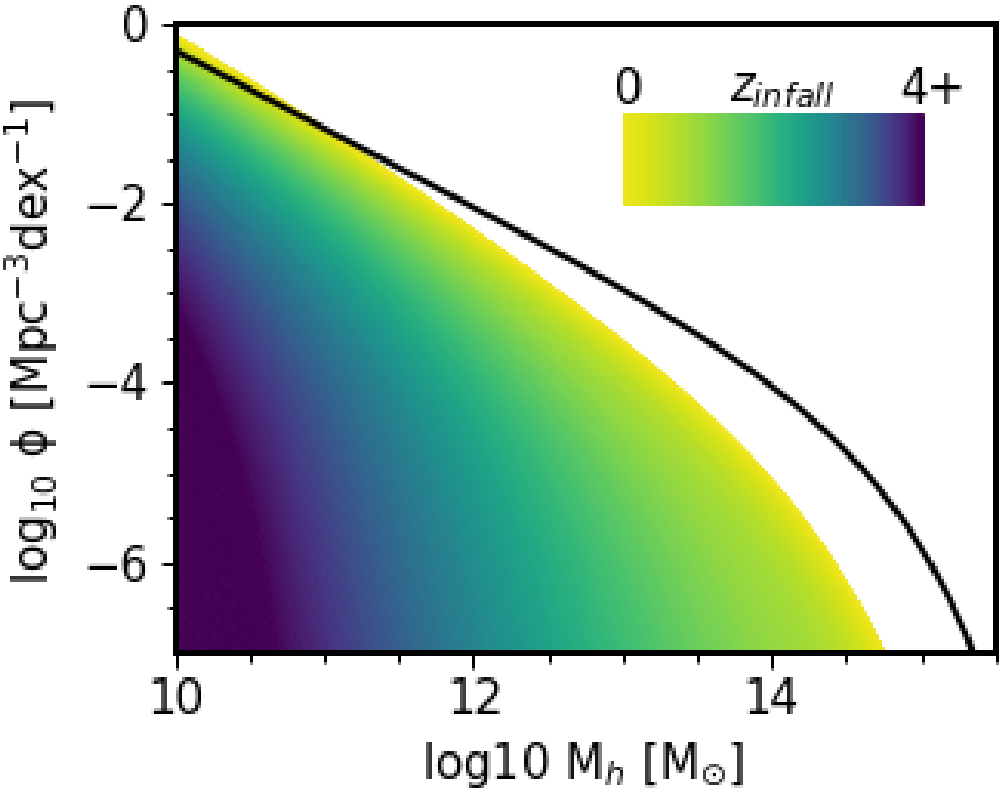


FIGURE 1.2: Central Halo Mass function, the global number density of haloes at redshift $z = 0.1$, black line. Global Unevolved Surviving Subhalo Mass Function at redshift $z = 0.1$, coloured section. The colour gradient shows the fraction of subhaloes at each mass coloured by the redshift of infall. Note due to the logarithmic scaling the thinner yellow bar at the top of the distribution is a majority contribution of subhaloes, showing that at all masses recent accretion dominates the population.

is assigned to be the larger of the two merging haloes[5]. At each merger a 'branch' is created following this branch which followed back in time gives the assembly history of that halo. We show an example of a simple merger tree in Figure 1.3.
There are increasingly more complex merger tree diagrams, as it is not the case that haloes merge simply we list below some examples of this complexity and show what these look like on a stylised merger tree in Figure 1.4:

- Substructure: A halo is not immediately dissolved upon accretion to the central halo, the substructure is therefore maintained in central haloes between epochs. Furthermore, haloes on accretion may also carry internal substructure, which creates 2nd (and above) order substructure in the parent halo.

[5]This remains true even if the main progenitor branch becomes smaller than the other branch at an earlier redshift.

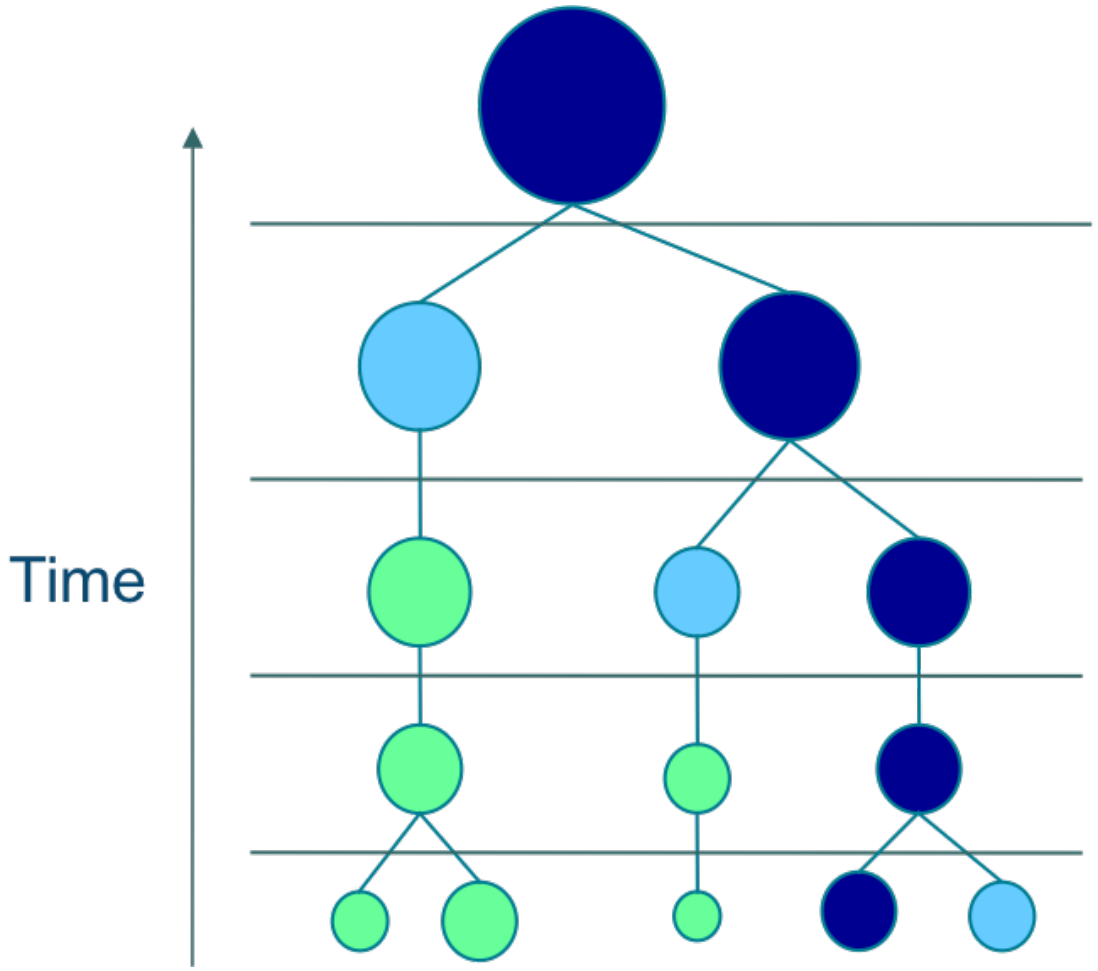


FIGURE 1.3: A cartoon of a simple merger tree. Increasing time moves up the page as indicated on the right, horizontal lines separate epochs. Dark blue circles represent the main progenitor or trunk haloes of the merger tree. Haloes that merge directly onto the trunk are coloured in light blue, the merger history or branch for each of these merging haloes is shown in green.

- Flyby Haloes: Some haloes may pass through another halo and may lose mass but are not captured. These haloes are known as flyby halos and must be accounted for when constructing mass functions to not inflate the number count of subhaloes.
- Re-Accreted/backsplash Haloes: Haloes on particularly eccentric orbits or haloes with initially high velocity may enter a halo but subsequently leave the halo group. These haloes, whilst still bound to the grou,p may not be associated with the group by a halo finder appearing to be flyby haloes. However, they at a later time these haloes may reenter the group: Firstly, one must ensure these halos are not double-counted. Secondly, caution must be taken that the mass they are accreted at is fit for purpose (i.e., mass at last accretion, mass at first accretion, or peak mass).

Whilst there is clearly a wealth of information available from dark matter simulations, they require a large amount of computational power both to run and analyse. For many

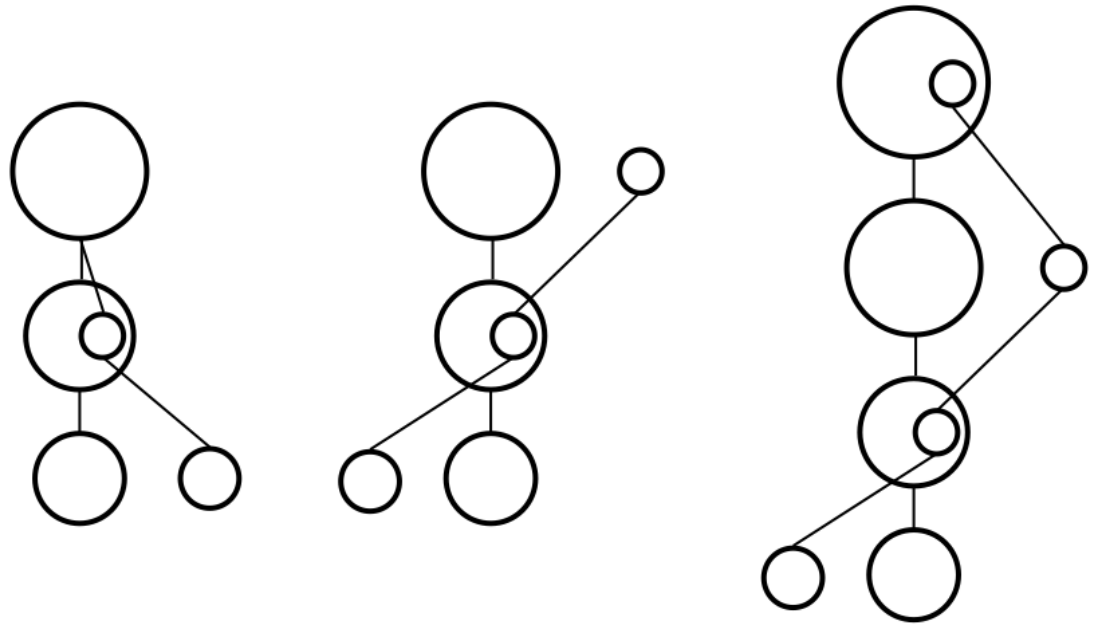


FIGURE 1.4: From left to right: An example of: 1^{st} order substructure, a flyby halo and, a re-accreted halo.

models of galaxy evolution the information provided is far in excess of what is required. Press-Schechter Formalism [48] provides an alternative way to generate simple analytic dark matter distributions at a greatly reduced computational cost by sacrificing spatial information. For a simple visualisation, one can consider the dark matter distribution at high redshift as a random Gaussian probability distribution.

Peaks in this distribution represent dark matter overdensities which will collapse over time where higher peaks collapse earlier due to greater gravitational influence. This can be visualised as a time-evolving threshold above which the dark matter density would qualify as a collapsed halo. If the summation of two Gaussian distributions are above this threshold then they become a single over-density or halo. A simplified cartoon is shown to visualise this in Figure 1.5.

Following Figure 1.5 considering the top dashed line as the earliest epoch we look for collapsed haloes by examining the dashed line; two peaks in the density field can be found above the threshold denoting two collapsed halo structures. The area above the threshold is analogous to the mass of the halo and the spread of the peak analogous to the size. At the second threshold, there is another collapsed peak mostly contributed by the brown Gaussian. Furthermore, the two peaks from the first collapse have grown in mass as can be seen by an increased area and spatial extent. Finally, looking at the last collapse threshold: On the left, there is a new isolated peak; on the right, the peaks created by the yellow and brown Gaussian distributions now create one continuous area above the threshold, this is an indication of two merged haloes. Within this merged halo, there are two distinct peaks that can indicate the nature of the halo substructure. It is

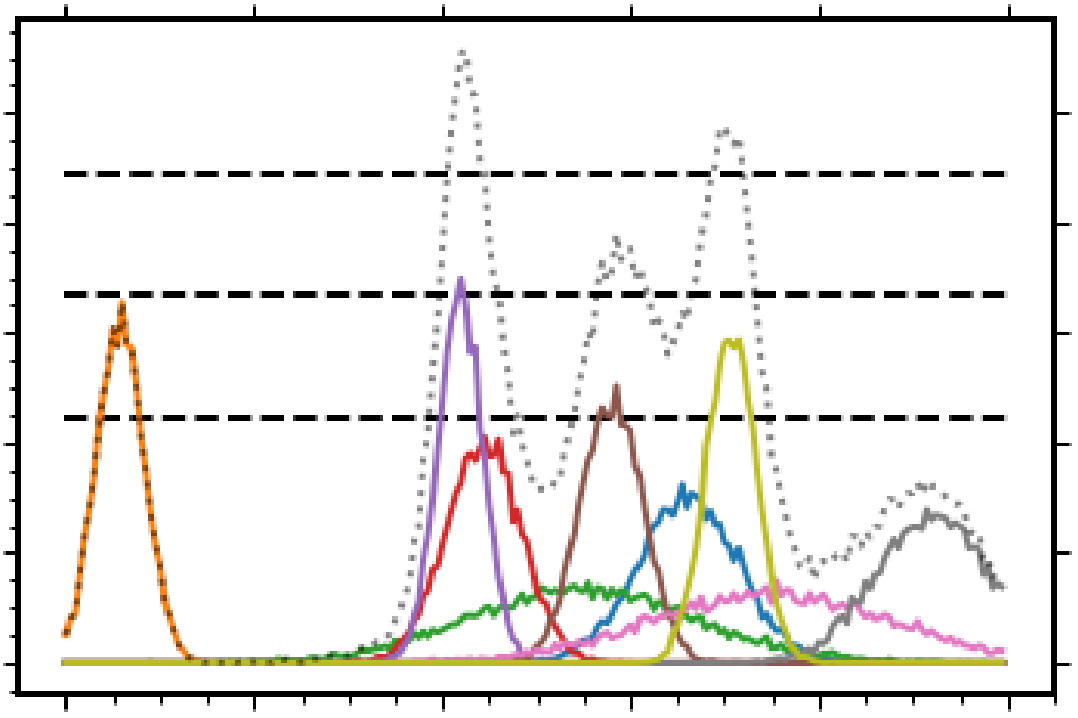


FIGURE 1.5: A visualisation of the core theory of Press-Schechter Formalism. The solid coloured lines are random Gaussian distributions representing fluctuations in the initial mass density field, the grey dotted line is the total mass distribution. The three black dashed lines are for visualising the threshold of collapse at different epochs. The x-axis is an arbitrary spatial dimension, and the y-axis an arbitrary density unit.

evident that at later epochs the merged substructures will also include the purple peak. This picture is an oversimplification of the Press-Schechter Formalism. For example, instead of simple collapse thresholds, a window function is used to smooth the density field. Finally, here the peaks are represented as static, however, their size and location will evolve with time. The shift in the location of the peaks can be used to build dynamic halo histories without the need for full N-body simulations [49].

The Press-Schechter Formalism can be used to analytically create a merger tree by utilising the probability of any two overdensities merging. An algorithm that starting at a given redshift with a given halo mass can then construct a theoretically valid merger history. We summarise here the GALFORM [50] algorithm for making merger trees from Press-Schechter Formalism[6].

- Pick a starting Mass, M_h and redshift z. This will be the final mass/redshift for the main progenitor halo in the tree.
- Step backwards in redshift in steps, dz, where dz is chosen that the probability, P, of a halo having a merger in dz is $P \ll 1$.

[6] In Parkinson et al. [51] improvements to this method are proposed to better align the generated merger trees with the millennium simulation by using a perturbing function to remove systematic differences between N-Body simulations and the algorithm.

- Generate a uniform random number, N, if $N < P$ the halo splits, else the mass is reduced to account for unresolved haloes or smooth accretion and the process repeats.

- If the halo has split a random number is generated from the probability distribution that defines the possible progenitor halo masses from Press-Schechter Formalism this is the mass of the split halo, M_{split}. The central halo mass is then updated to be decreased equal to the split haloes' mass and a smooth/unresolved accretion component and the process repeated.

The merger trees generated using this algorithm are simpler and less computationally expensive than those extracted from N-body simulations, and if calibrated properly they are fully consistent with ΛCDM cosmology. They, therefore, are an ideal choice for simulations that need to generate halo merger trees on demand.

1.3 Hydrodynamic and N-body Simulations

Hydrodynamical simulations are the most physical galaxy models simultaneously simulating baryons and dark matter. Within a volume dark matter, gas, and stars are assigned to particles and/or grid cells. Forces between particles are then applied to each element with respect to other elements. The volume is then advanced by repeated numerical solving and time stepping. At set intervals, snapshots are produced to be used for analysis. Processes that take place below the resolution limit of the simulation, such as star formation, are simulated using sub-grid analytic routines based on the properties of the given volume element.

1.3.1 State of the art: Illustris TNG

The Illustris TNG simulations are a suite of 18 simulations, which use three different volumes, $\sim 50, 100, 300\ Mpc^3$, and include a range of physical processes and resolutions. TNG at its core uses the code AREPO [52], a mesh-based code, which solves coupled ideal magneto-hydrodynamics and self-gravity using periodic boundary conditions. Even for TNG50, where exceptionally high resolution is achieved, much of the baryonic physics falls below the resolution limit, thus relying on 'sub-grid' models. Sub-Grid models are used for star formation, feeding of supermassive black holes, galaxy feedback process (star formation and AGN), and more. A simple specific example of one such sub-grid model is star formation: Gas above a threshold density of $n_H \simeq 0.1 cm^{-3}$ is allowed to form stars following the numerical results of Springel and Hernquist [53]. Sub-grid models attempt to replicate results of smaller higher resolution simulations or from observations.

1.3.2 Pros and Cons

The drawbacks of using a hydrodynamic model are foremost limitations of resources. They require vast amounts of computational resource to run and take a lot of developer time and skill to create. Furthermore, they create huge amounts of output data that can be difficult to process and interpret. Last but not least all simulations are subject to volume and resolution constraints. For example, the TNG simulations run three box sizes; the larger boxes can simulate rare/low abundance objects, whereas, the smaller boxes can simulate galaxies in high detail. The computational time to run a simulation that produces both high detail and rare objects would make the computational cost impractical and the data products unmanageable with current tools and technologies.
The advantages of explicitly modelling galaxies using hydrodynamics is that one can study the dynamics and interactions of the particles pertaining to each galaxy or dark matter halo directly. Furthermore, by carefully choosing the size and resolution of the simulation, it is possible to probe the connection between large and small scale structures over many orders of magnitude. For example, in TNG it is possible to explore how the cosmological environment affects galaxy formation, the gas flow along dark matter filaments and, the statistics of galaxy mergers. Alternately in FIRE [54], an advanced smooth particle hydrodynamic simulation, one high resolution 'zoom in' region is simulated with high fidelity. The particle resolution in FIRE is such that feedback from star formation winds, AGN, and even individual supernovae are resolved. Hydrodynamics represent the state of the art in our ability to simulate a universe or an individual galaxy holistically. However, the high computational and human cost associated with their running and development means they are not the ideal tool for rapid testing of new theories or ideas.
Progress has been made in each aspect that limits hydrodynamics. Firstly, the speed, memory and architecture of computing resources continue to improve with recent advances in GPU technology contributing to the overall power available to simulators. In addition to this, novel techniques have been developed such as "genetic modification", whereby the initial conditions of a simulation are subtly altered to test how perturbations and modelling assumptions affect self-similar galaxies [55]. Building on these advances in speed and flexibility it may well become possible to use hydrodynamic simulations as comprehensive tools. At present we must still utilise diverse sets of complementary tools available to deal with the highly non-linear and complex processes regulating galaxy formation and evolution.

1.4 Semi-Analytic Modelling

1.4.1 History of Semi-Analytic Models

Building off the hierarchical collapse models from ΛCDM cosmology and the power of generating merger trees from extended Press Schechter (EPS) routines [48], a new form of galaxy model was conceived named semi-analytic models. In a dark matter-dominated universe, galaxies must be influenced by the structure formation of the dark matter haloes predicted by ΛCDM cosmology. The earliest models successfully recreated the galaxy luminosity function by assigning the baryonic matter to EPS haloes and solving analytic gas collapse equations [56].

Semi-analytic models were then developed and extended to include more advanced galaxy physics, such as the formation of galaxy disks [29]. Increasing in complexity by both folding in more advanced physics such as AGN feedback and using merger trees from N-body dark matter simulations, the models were able to increase the scope of predictions by tuning parameters over several simulation runs to best fit the observations [57]. Continuing improvements in cosmological (dark matter) simulations have allowed for semi-analytic models to improve both in resolution and size. In addition, as computing power has become more accessible and improved Monte Carlo methods have been developed, this has allowed for the development of a variety of multi-parameter semi-analytic models tuned over hundreds of thousands of runs [1, 18, 31, 58, 59]. In this section we will discuss the state of the art in semi-analytic modelling by summarising a review by Somerville et al. [1], followed by deconstructing a subset of the latest models, and finally, present a consideration of the drawbacks of semi-analytic models compared to other galaxy modelling techniques.

1.4.1.1 Somerville (2015) Review [1]

The semi-analytic models used here have been described in detail in Somerville & Primack (1999), Somerville et al. (2001) and most recently in Somerville et al. (2008a, hereafter S08) and Somerville et al. (2012, S12). The Santa Cruz modelling framework has also recently been described in Porter et al. (2014). We refer the reader to those papers for details.

This review explores three models published in Somerville et al. [22] (hereafter S08), Somerville et al. [60] (hereafter S12), Porter et al. [61] (hereafter Santa Cruz). Each of these models is run using EPS merger trees as a backbone. Using analytic merger trees allows for high-resolution simulation reducing the minimum halo size. Reducing the halo size, in turn, allows the simulation of smaller galaxies. Galaxies grow via the cooling of gas from re-ionisation. Before this point haloes are seeded with gas at the cosmic baryon

fraction. After this gas cools as given in [62], it condenses and forms stars at the centre of dark matter haloes.

The prescriptions made to handle the hierarchical assembly predicted by ΛCDM cosmology concern firstly the halo structures. At the point of halo merger, the larger halo and associated galaxy are assigned as the central, the smaller halo(es) and associated galaxies become subhaloes and satellite galaxies. Satellite galaxies then merge with the central galaxies on timescales given by the Chandrasekhar formula [63], as given in, for example, Boylan-Kolchin et al. [e.g. 64]. The mergers of satellites are one of the methods that transform the morphologies of central galaxies, mergers gradually reduce the central galaxies angular momentum eventually creating a spheroid [65].

In addition to changing the morphologies of central galaxies, mergers are also thought to drive one of the two pathways of star formation. During a merger the gas of the two merging galaxies is destabilised triggering fragmentation and a galaxy-wide "starburst", a short period where star formation is enhanced by up to an order of magnitude [65, 66]. The second mode of star formation is referred to as "disk mode" and is usually modelled in three ways: The Kennicutt-Schmitt law [67] assumes that the surface density of star formation, Σ_{SFR}, is related to the surface density of cold neutral gas, Σ_{gas},

$$\Sigma_{SFR} = A_{SF}\Sigma_{gas}^{N_{SF}}. \tag{1.1}$$

It is here and in the following star formation rate recipes that the first tuning parameters are found; A_{SF} and N_{SF} control respectively the normalisation and slope of the star formation rate law. In addition, there are two critical gas densities Σ_{crit} and $\Sigma_{H2,crit}$, below which neutral gas or molecular gas will not form stars. Similar to the Kennicutt-Schmitt recipe, molecular gas star formation follows empirical results that link the molecular gas surface density to the star formation rate as empirically determined by Bigiel et al. [68],

$$\Sigma_{\mathrm{SFR}} = \left(\frac{A_{\mathrm{SF}}}{10 M_{\odot}\mathrm{PC}^{-2}}\right)\Sigma_{\mathrm{H_2}} N_{\mathrm{SF}}. \tag{1.2}$$

It's observed that above a critical H2 density the SFR steepens [69], inspired by this the SFR may also be modelled using a two part scaling law,

$$\Sigma_{\mathrm{SFR}} = A_{\mathrm{SF}}\left(\frac{\Sigma_{\mathrm{H_2}}}{10 M_{\odot}\mathrm{pc}^{-2}}\right)\left(1+\frac{\Sigma_{H_2}}{\Sigma_{\mathrm{H_2,crit}}}\right)^{N_{\mathrm{SP}}}. \tag{1.3}$$

Semi-analytic models rely on several feedback prescriptions to regulate their star formation. The energy released during supernovae, created as a result of star formation is deposited in the inter stellar medium consequently driving outflows of cold gas at the rate,

$$\dot{m}_{\mathrm{out}} = \epsilon_{\mathrm{SN}}\left(\frac{V_0}{V_c}\right)^{\alpha_{\mathrm{rh}}}\dot{m}_{*}, \tag{1.4}$$

where $\epsilon_{\rm SN}$ is a parameter that scales the efficiency of the supernovae's ability to drive outflows, and the remaining parameters control the binding properties of the gas. In addition to supernova feedback, most SAMs include AGN feedback caused by energy/-momentum release from the supermassive black hole at the centre of the galaxy. AGN feedback is believed to act in two distinct modes: The first mode, "quasar mode", is initiated after a galaxy merger, as the black hole grows in mass, more energy is released to the ISM until the gas accretion is halted by outflows from the black hole. In this mode, AGN feedback gradually reduces the SFR by heating the cold gas reservoir. The second AGN feedback mode, "radio mode", is commonly assumed to be in the form of a jet and triggered when accretion drops below a few per cent Eddington. This feedback can also prevent late cooling gas in the central regions of galaxies [70].
The takeaway of this brief introduction to semi-analytic modelling is that they can be characterised by heavy parametrisation of each singular observable which results in a very highly parameterised model. This is due to the large uncertainties in the choice (and strength thereof) of the physical processes to include. The interested reader can find additional semi-analytic methods that control the properties of the gas described in Appendix A. The following subsections will describe two state of the art semi-analytic models.

1.4.2 State of the art

Many of the modern semi-analytic models were conceived in the mid-two-thousands and have been iterated on for over a decade, due to increasing complexity without significantly improving convergence in results thus they are arguably reaching a plateau in usefulness. During this time the amount, quality and availability of comparison data have significantly increased. As the models have matured they have been continually added to with many physical prescriptions to match a wide variety of observations. Despite this, several models still struggle to reproduce the basics of galaxy evolution such as the evolution of the stellar mass function [71], or it can be found that different flavours of model can produce disparate results [72]. Here we briefly discuss two models that represent state-of-the-art in semi-analytic galaxy modelling.

1.4.2.1 GAEA

The GAEA model is a good example of a mature semi-analytic model, originating from the seminal work modelling assembly of BCGs by De Lucia and Blaizot [73]. It has subsequently been updated many times in 2008 (De Lucia & Helmi) 2010 (Li et al.) 2014 (De Lucia et al.) and Hirschmann et al. [74] (see references within). One of the latest renditions of this model based on that of [75] and updated in [76], this version

of the model includes a description of quasar driven winds in an effort to reproduce the size-mass relation for both early and late-type galaxies.

1.4.2.2 SHARK

The SHARK semi-analytic model represents an attempt to keep the field of semi-analytic models in line with current software engineering best practice. The release paper [77] carefully details the physical prescriptions and modularity that defines the code. SHARK is designed to be flexible and modular and takes pride in being the most accessible open-source code available. In Lagos et al. [77] the authors show the baseline performance of SHARK to be as good as or exceeding other semi-analytic models in the mass-size, gas-stellar mass and stellar mass-metallicity relations. The flexibility of SHARK is then demonstrated in Lagos et al. [78], where SHARK is modified to use different dust mass models to match spectral energy distribution observations of galaxy emission from the far-UV to the far-IR.

1.4.3 Drawbacks of Semi-Analytic Modelling

A major feature of Semi-Analytic modelling is the number of parameters that can be tuned to reproduce observations. For example, the default SHARK configuration has over 50 input parameters included[7]. Whilst the breadth of these parameters allows for semi-analytic models to produce fits to a wide range of observational properties, they also inevitably present serious degeneracies, thus obscuring the actual significance of the specific physical processes assumed in input [79, 80].

1.5 Semi-Empirical Modelling

Semi-empirical models similarly to semi-analytic models are often built on numerical or analytical dark matter merger trees [33]. A semi-empirical model uses a set of observations as inputs to build the model upon and another set of observations to constrain the output. For example, many semi-empirical models us the stellar mass function to calibrate the mass and mass growth of galaxies. In brief, this is achieved by comparing the relative abundances of dark matter haloes and observed galaxies. Under the assumption that galaxies reside in parent haloes that share a similar number density, an analytic correlation between halo mass and stellar mass is determined [81, 82]. This process is called "abundance matching" and is explained in greater depth in Section 2.4.2. When used correctly, abundance matching ensures that semi-empirical models recreate the observed distributions of galaxies over many redshift epochs. By using the growth of dark matter haloes, merger trees, and abundance matching the need for modelling of galaxy mass and

[7]Many of there parameters can be highly constrained by priors and the following applies to where non orthogonal parameters are not well constrained

growth is removed via empirical constraints. Constraining galaxy masses and/or other properties via observations significantly reduces the modelling parameter space compared to hydro-dynamical or semi-analytic models. A reduced parameter space increases the transparency of the model by making the effects of the other parameters more clear. By empirically fixing properties that may not yet have clear constraint or physical mechanism, or those properties that are not of interest to the current simulation, semi-empirical models are able to focus on the reproduction of a singular or small subset of properties. Since their inception, semi-empirical models have grown in complexity and in doing so have naturally increased the number of modelling parameters. As the complexity of the models is controlled by observational priors and the model should be *internally consistent*, and thus it is possible to accurately probe very specific aspects of galaxy formation. It is, however, possible to create SEM where the observational priors, and by extension the model, are *internally inconsistent*, i.e. one or more of the observational inputs do not agree. An example of such inconsistency can be found in the observed stellar mass function and the observed star formation rate. The time integral of the star formation rate predicts far more stellar mass growth than is predicted by the observed growth of the stellar mass function [83, 84]. If these two empirical properties were included in a semi-empirical model then the model would be *internally inconsistent*. However, the semi-empirical model could use the stellar mass function as an input to show the implied star formation rate without needing the conflicting measurement to form a physical part of the model. Similar yet unappreciated effects involving systematics in galaxy assembly make up a significant part of Chapters 4 & 5.

1.5.1 State of the art

Here we describe three state-of-the-art semi-empirical models. Each of the models described here has a radically different approach to semi-empirical modelling. They are each constrained by observational priors and make predictions about the buildup of mass in galaxies through mergers and star formation rate. The range of approaches show the flexibility and the agreement between models show the power of the empirical approach.

1.5.1.1 Rodríguez-Puebla et al. [2]

In the model presented in Rodríguez-Puebla et al. [2], the SMHM is constrained using the 2-point correlation function and the evolution of the SMF. Using these constraints to predict the growth of galaxies instead of using SFR the inconsistency between observed SFR and SMF evolution is avoided. The main result of this paper is valuable constraints on the SFR and quenched fraction of galaxies in multi-parameter space, in both halo mass - redshift and stellar mass - redshift as shown in Figure 1.6.

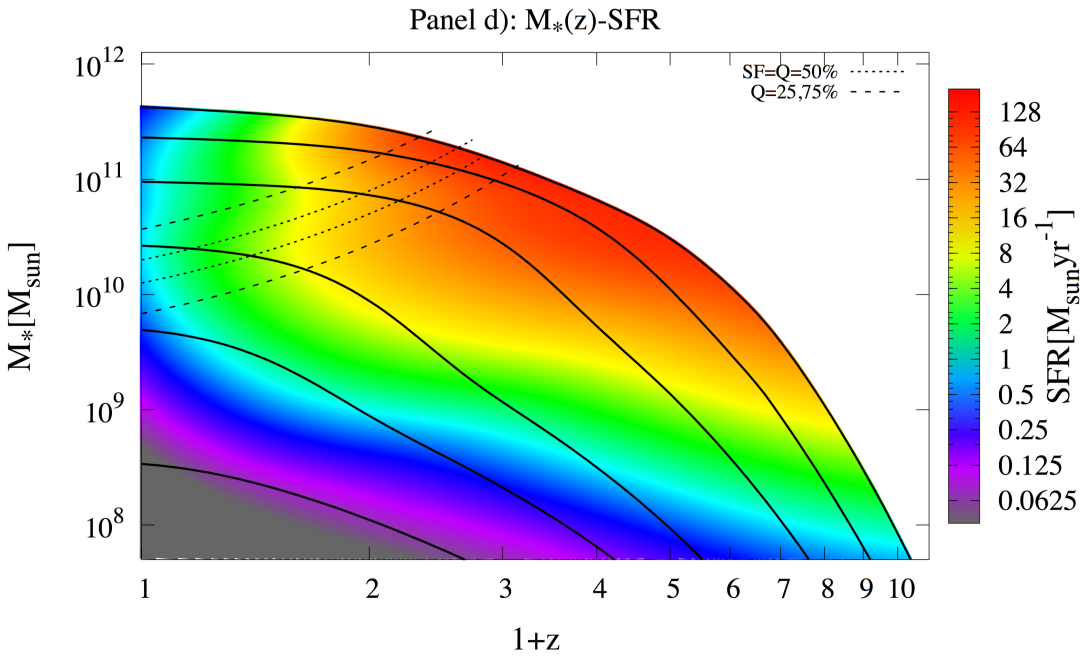


FIGURE 1.6: In this figure, the relationship between mass redshift and the star formation rate is shown. The black lines show the 'galaxy mass tracks', the average path of the galaxies through M_*-z space. The colour shows the star formation rate associated with an average galaxy at that point in M_*-z space. The dotted and dashed lines show the location where a percentage of galaxies (as labelled) have been quenched. Image is used with permission from: Aldo Rodriguez Puebla [2]

The key message from Figure 1.6 is that most stellar mass is built up in place at high redshift as the mass growth below redshift $z < 1$ is a few per cent. In Rodríguez-Puebla et al. [2] it is also shown that massive galaxies then quench at the point where the sSFR is equal to the specific halo mass accretion rate which happens at $M_{vir} \sim 2 \times 10^{11} M_{\odot}$, this point is shown on the figure as the dotted/dashed lines where it can be seen that the SFR rapidly decreases.

1.5.1.2 EMERGE

EMERGE, presented in Moster et al. [17], builds upon N-body dark matter simulations. From the dark matter simulation each halo is tracked with its growth history and relation to other haloes. EMERGE works under the core assumption that galaxies grow in conjunction with the dark matter haloes they reside in. This is achieved through coupling the star formation rate of each galaxy to the accretion rate of the host dark matter halo parameterised in the following way,

$$\begin{aligned}
\log_{10} M_1(z) &= M_0 + M_z(1-a) = M_0 + M_z\frac{z}{z+1}, \\
\epsilon_{\rm N}(z) &= \epsilon_0 + \epsilon_z(1-a) = \epsilon_0 + \epsilon_z\frac{z}{z+1}, \\
\beta(z) &= \beta_0 + \beta_z(1-a) = \beta_0 + \beta_z\frac{z}{z+1}, \\
\gamma(z) &= \gamma_0,
\end{aligned} \tag{1.5}$$

where M, $\epsilon_{\rm N}$, β and γ are respectively the characteristic mass, efficiency normalisation, low mass slope and high mass slope. Each parameter has a value at redshift zero given by, e.g. M_0, and, with the exception of γ, a redshift evolution coefficient given by, e.g. M_0. Additional parameters are included controlling for example; the point at which satellite galaxies are disrupted returning all mass to the ICM, the fraction of mass ejected to the ICM during a merger, and quenching parameters from Wetzel et al. [9], e.t.c. The fractions of 'in-situ' vs 'ex-situ' mass build up are computed and fit by,

$$f_{\rm acc}(z) = f_2 \exp\left[-f_1(z+1)\right] \tag{1.6}$$

as well as the star formation histories that are fit by,

$$\log \Psi(z) = -\log\left[\Psi_1(z+1)^{-\Psi_2} + {\rm e}^{\Psi_3(z+1)-\Psi_4}\right] \tag{1.7}$$

.

Similarly to Rodríguez-Puebla et al. [2] it is found that stellar mass is built up in-situ until a breakpoint at $1.1 \times 10^{12} M_\odot$. Above this point, larger galaxies may then accrete a significant proportion of their mass.

1.5.1.3 UniverseMachine

UniverseMachine [6] is built on a background of halo merger trees extracted from the Bolshoi simulation [41, 85], using the ROCKSTAR halo finder and the Consistent Trees codes [44, 86]. There are 21 modelled outputs with 44 tuning parameters and 6 priors. Using a Markov Chain Monte Carlo the parameters are fit to a large set of observations though the algorithm depicted in Figure 1.7.
UniverseMachine works by iterating over three key steps:

- Parameter selection and probability distribution generation. This step phase samples a subsection of the parameter space and computes the associated star formation probabilities that exist within it. (Fig. 1.7: 1 & 2)

- Mapping into and integration along halo merger trees. The SFR distributions are mapped into haloes and integrated forward in time along the halo merger trees to estimate central galaxy masses. (Fig. 1.7: 3 & 4)

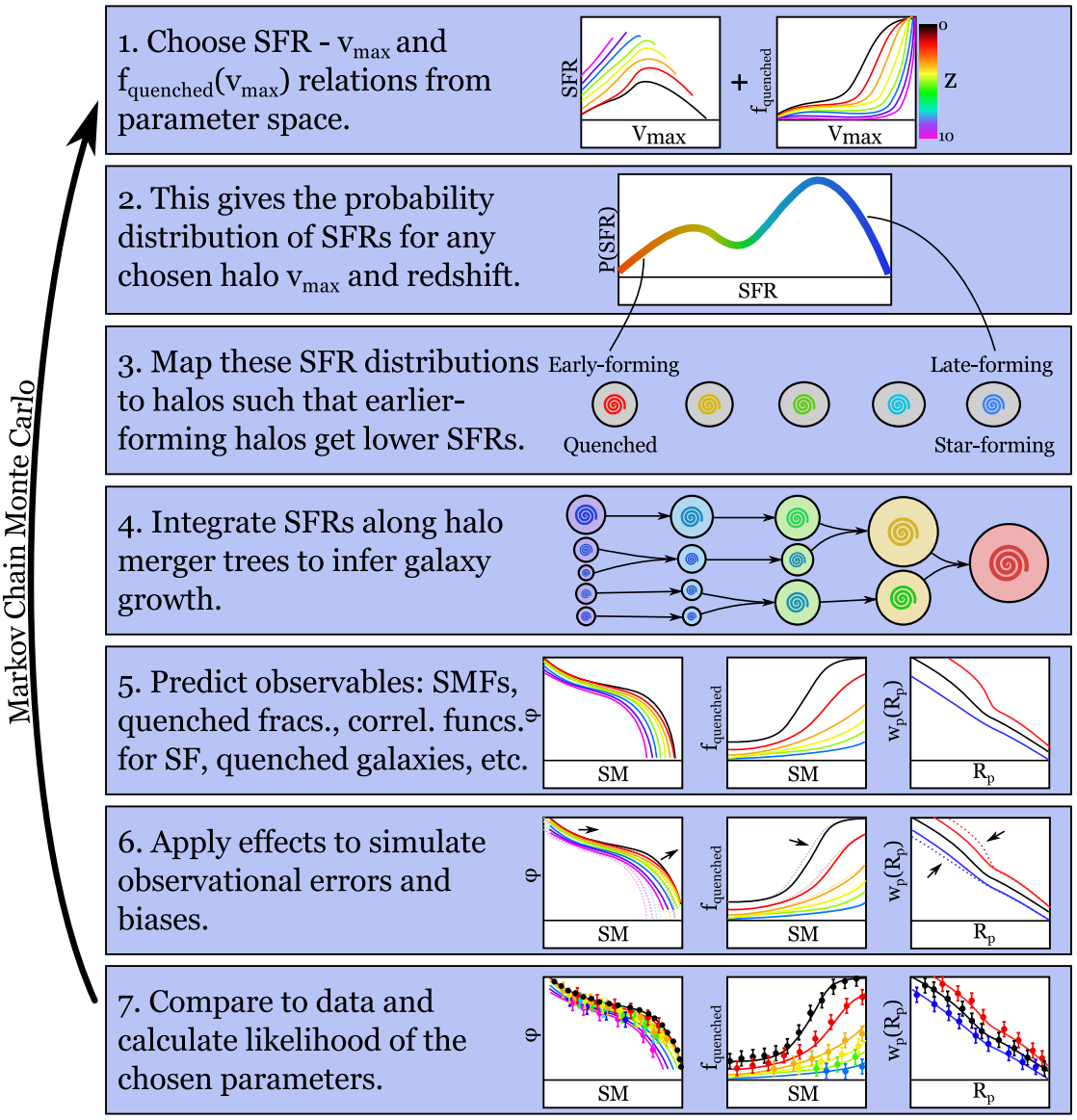


FIGURE 1.7: This figure shows the iterative process used by the Markov Chain Monte Carlo in UNIVERSEMACHINE. As described by the labels in the figure relations from the empirical parameter space are chosen (1), giving a SFR distribution (2). The distribution is mapped into haloes (3), the SFR is integrated along the merger histories to infer the galaxy growth (4). Observable features are produced (5), and observational bias corrections applied (6). Finally, this is compared to the data and the likely-hood of the parameters produced and new parameters picked accordingly. Image used with permission from: Peter Behroozi [6]

- Observable simulation. A set of observables are then extracted from the galaxy population and given simulated observational errors. These are then compared to data and the likelihood of the parameters chosen in the first step calculated. (Fig. 1.7: 5, 6, & 7)

This model is massively parallel allowing for over 10^5 cores involved in any given parameter estimation. The model has parametrisation levels approaching that of a Semi-Analytic model however produces numerous predictions included in which are 'in-situ'

vs 'ex-situ' growth ratios which are in broad agreement with other models.

1.5.2 Limitations of the empirical technique

Semi-empirical models, as demonstrated above, come in many different flavours. Whilst data is what empowers semi-empirical models, it also presents one of semi-empirical models major limitations. Having wide and deep data sets which give statistically representative samples via large volumes to high redshift is difficult and costly. Most techniques involve correcting data sets and surveys to be consistent with one another creating a patchwork connecting the local and distant Universe. Semi-empirical models will continue to improve along with the available data. However, when built upon traditional dark matter simulations they will share the volume constraints found in semi-analytic and hydro-dynamical models.

1.6 Direct Model Comparison

In this chapter we compare several different modelling techniques and discuss specific examples in terms of volume, resolution and number of modelling parameters. In Table 1.1 we quantitatively show the differences between STEEL [19–21], Rodríguez-Puebla et al. [2], UNIVERSE MACHINE [6], Somerville et al. [22][8], and TNG [23]. As described in the prior sections of this Chapter each 'flavour' of galaxy model uses/interprets volume, resolution, and modelling parameters differently. So in many cases the direct comparisons between quantitative descriptors must be caveated by considering the models in question.

1.7 Galaxy Surveys Present and Future

Galaxy surveys are the cornerstone of semi-empirical galaxy modelling. Combined with dark matter simulations they form the basis of abundance matching. Other properties such as star formation, colour, shape, size, position, e.t.c., can all form constraints for, either inputs or outputs, of semi-empirical models. It therefore stands that improving the quality of models is predicated on improving the quality of the data.

1.7.1 Past and Ongoing Surveys

The galaxy surveys essential to semi-empirical models come in two forms 'wide' and 'deep'. Wide surveys look to cover a large area of the sky to get a statistically significant sample of the galaxy population. Deep surveys cover a much smaller area but use longer

[8] To determine the 'true' number of prameters it is required to communicate directly with the researchers who built or worked on the model. Many parameters may not be reported as they are 'hard coded', thus to determine these numbers the authors of each model were contacted but unfortunately not all were able to respond.

Model	Method	Volume Mpc^3	DM Halo Resolution $M_\odot$	Parameters (free) (constrained)
STEEL	Statistical SEM	N/A (Statistical)	Arbitrary	15-30 (1-5) (14-29)
R-P+2017	SEM	250 - 1000 h^{-1}	9.6×10^7 - 8.7×10^9	18 (18) (0)
UniM	SEM	250 h^{-1}	$1.6 \times 10^8 h^{-1}$	46 (44) (2)
Somerville+08	SAM	N/A (Press Shecter e.g. Parkinson et al. [51])	10^{10} (10^{11} smallest root)	No response see Footnote 8
TNG (50, 100, 300)	Numerical	51.7, 110.7, 302.6	4.5×10^5, 7.5×10^6, 5.9×10^7	135 (102) (33)

TABLE 1.1: A summary of different codes [2, 6, 19–23] in terms of method, volume, resolution and modelling parameters.

exposures to observe much fainter objects. The trade-off between width and depth is visualised by a simple example in Figure 1.8. Tight surveys collect more photons per unit area and can, therefore, see to a greater depth.

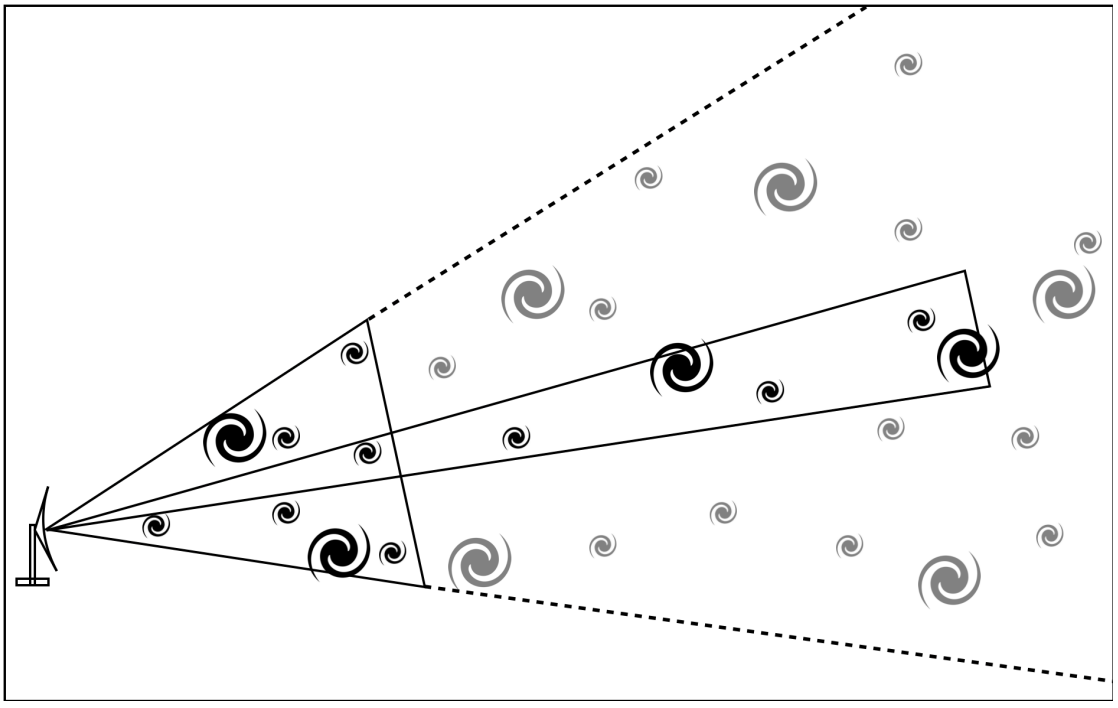


FIGURE 1.8: A simplified visualisation of how a trade-off is made between width and depth in surveys. Given two observation areas, the telescope is able to pick up galaxies to a much greater depth in narrow observation. By spending more time on one area of the sky the telescope can pick up more photons and therefore observe to a greater depth.

In addition, larger galaxies are significantly brighter (the example in Figure 1.9 uses two

galaxy size examples). The smaller galaxies can be observed to the solid line and the larger galaxies to the dashed line. The telescope observes 6 small galaxies and 5 large galaxies, however as the larger galaxies are brighter and can be seen to a greater depth. Corrections are therefore made to the number density of galaxies based on the depth that the galaxy can be observed to. In the 2D example from Figure 1.9 the small galaxies can be seen in an area given by the solid area which we can call 1 unit^2 the larger galaxies can be seen in an area given by the dashed area that is 4 unit^2. In this example, the smaller galaxies have a number density of 6 galaxies p.u. area, the larger galaxies 1.25 galaxies p.u. area.

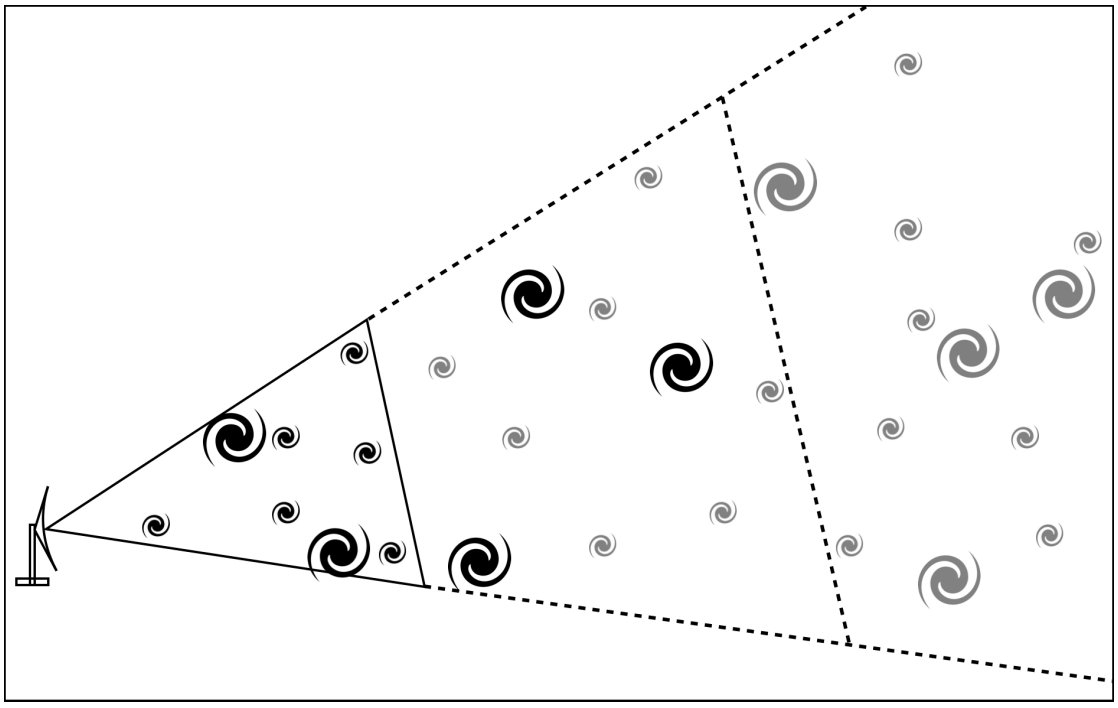


FIGURE 1.9: A simplified visualisation of how the brightness of galaxies affects the observation depth. The smaller, dimmer galaxies can be seen in the solid cone, larger and brighter galaxies can be seen up to the limits of the dashed cone.

For each galaxy luminosity observed in a survey, a correction must be made for the total depth it could have been observed to and the associated volume. In this way, the volumetric density correction can be calculated for each galaxy in a given survey.

1.7.1.1 Sloan Digital Sky Survey (SDSS)

The Sloan Digital Sky Survey (SDSS) is arguably the premier wide survey for galaxies. SDSS started in 1998 with the mission to catalogue and map as much of the universe as possible. The latest release of the fourth phase is SDSS-DR19 [87] now includes one-third of the night sky, it covers (u,g,r,i,z) wavelengths, has over 1.2 billion objects including over 200 million galaxies.

1.7.1.2 CANDELS

CANDELS (The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey) observed galaxies in the redshift range z = 8 to 1.5. CANDELS uses three Hubble Space Telescope (HST) cameras to capture emission from mid-UV to near-IR for 250,000 galaxies. The primary goals of this survey of interest to this thesis is the observation of the peak of star formation and AGN activity at $z \simeq 2$ [88].

1.7.1.3 COSMOS

COSMOS: The Cosmic Evolution Survey is a 2 square degree survey covering many wavelengths using a combination of observation from many space-based and ground-based telescopes (Space: Hubble, Spitzer, GALEX, XMM, Chandra, Herschel, NuStar) (Ground: Keck, Subaru, VLA, ESO-VLT, UKIRT, NOAO, Badde and Blanco, CFHT). Combining the data allows COSMOS to create a rich view of the observed objects across a huge range of wavelengths. COSMOS was designed to observe galaxy evolution across cosmic time and understand the environmental effects of galaxies growing in groups and clusters [89].

1.7.1.4 3D-HST

Using 248 orbits and 124 pointings the 3D-HST is a highly complementary survey observing 625 arcminutes of previously observed extra-galactic fields [90]. 3D-HST provides spectroscopic redshifts for over 10,000 galaxies in the range $1 < z < 3.5$. The survey provides 22 bands of rest-frame colours used for the redshift catalogues, the UV-IR star formation rates and, stellar masses. The primary science goal is to investigate the evolution of galaxies at high redshifts and in combination with CANDELS be the primary spectroscopic data set until the launch of JWST.

1.7.2 Future Surveys

Due to the pace of technological advancements and the difficult and expensive nature of building (and launching in the case of space-based missions) telescopes, our current survey telescopes are far behind what is theoretically possible. In this subsection two of the next generation survey telescopes are described. Both are space-based missions being launched into orbit at the second Lagrange[9] point 4 times the distance between the moon and the earth. The second Lagrange point is unique in its position as it sits behind the earth and is therefore permanently in the earth's shadow. This positioning is ideal for satellites as it is also within the earth's magnetotail and therefore experiences

[9]The Lagrange points are 5 distinct points in a gravitational two-body system at which an object can remain at rest relative to the two major bodies. In these points, the forces of gravitational acceleration, centripetal acceleration and for some points the Coriolis effect are all in balance.

lower solar wind irradiation. The drawback of this point is that from here any faults that are present at launch or develop during the operation of the missions will be permanent as it is not possible to retrieve and/or fix objects sent to these points.

1.7.2.1 JWST [3]

The James Webb Space Telescope is a cold infrared telescope with a planned launch in 2021[10]. Using multi-object near-infrared spectroscopy JWST will be able to take spectra for 100 galaxies simultaneously. Galaxies will be observed in the redshift range $1 < z < 7$ covering the peak of cosmic star formation and much of their early formation. JWST is an infrared survey is particularly suited to probe the star formation of galaxies, the buildup of the Hubble sequence, the formation of metals, or the effects of starbursts [91].

1.7.2.2 EUCLID

The EUCLID satellite will carry two instruments VIS & NISP. Between them they will cover from 500nm to 2000nm wavelengths down to the 24th magnitude, providing images of galaxies up to redshift $z = 2$. Euclid will survey 15,000 deg^2 of the sky and then perform a deep survey covering 40 deg^2 to 2 magnitudes deeper. The primary mission goals of Euclid are to address open questions on dark matter, dark energy, cosmic expansion and the formation of large scale structures [92]. To address these questions galaxies will be used as tracers of dark matter that providing tests of cosmological models. To achieve this goal the EUCLID pipeline will contain advanced galaxy cluster detection algorithms [93].

1.8 STEEL

There are several outstanding issues in galactic astrophysics that the models and techniques described in this section are not able to comprehensively address. In this thesis, we will describe the development of STEEL as a tool to be able to directly address these issues by introducing a new modelling paradigm.

Firstly, we address the reproduction and evolution of the full stellar mass function history, including satellite and central galaxies over a wide range of masses. This has proven problematic for semi-analytic models [71] and is essential to understand how galaxies have grown over cosmic time. The importance of the full SMF reproduction is highlighted in Chapter 2 where we show how STEEL as an empirical model reproduces this by design. The generation of full satellite galaxy distributions is shown in Chapter 3 and compared to the distributions of satellites locally and at high redshift.

[10] Correct at time of writing, JWST has some notoriety for pushing this date back

Secondly, in-situ (internal mass growth) vs ex-situ (mass growth from cannibalisation of other galaxies) invites several open questions: The total amount from each source, the characteristic mass at which galaxies switch from one mode to the other, e.t.c. [6, 17, 94–96, e.g.]. We use the empirically generated/constrained environments from Chapter 3 to add constraints to these questions in Chapter 4.
Additionally, the selection of galaxy pair fractions and their evolution with redshift is in contention (Mundy et al. [4], Man et al. [97]). It is unclear if the pair fraction increases or decreases with redshift, which has implications for our ability to observationally estimate the galaxy merger rate. In Chapter 5 we show a systematic analysis of how the choice of stellar-mass estimation propagates into observed pair fractions.
Finally, the rate and significance of galaxy-galaxy mergers as predicted by ΛCDM hierarchical assembly is a source of debate [e.g. 32, 65, 98–100]. A core prediction of mergers is that they drive the morphological change seen in galaxies between high and low redshift [e.g. 101–103]. To conclude, in Chapter 5 we use the statistical merger rate obtained from STEEL alongside a statistical representation of several morphological change models that are driven by mergers. The results of these models are then compared to the morphological fractions in the local universe.

1.9 Original Sources

A large fraction of this thesis is a restructuring of three papers published during PhD studies by PJG. The papers are:

- *'A Statistical Semi-Empirical Model: Satellite galaxies in Groups and Clusters'* Grylls et al. [19], hereafter Paper I
- *'Predicting fully self-consistent satellite richness, galaxy growth and starformation rates from the STastical sEmi-Empirical modeL* STEEL*.'* Grylls et al. [20], hereafter Paper II
- *'The significant effects of stellar mass estimation on galaxy pair fractions.'* Grylls et al. [21], hereafter Paper III

Each chapter, including the introduction and conclusion, may contain work from all three sources however the major sources for each are as follows: Chapter 2 reports primarily from Paper I where the method was originally published, updates to the method from Paper II & Paper III are also included. Chapter 3 restructures the discussion of galaxy distributions from Paper I & Paper II. Chapter 4 presents the critical finding published in Paper II and contextualises the main result in greater detail. Finally, Chapter 5 details the results from Paper III and a work in preparation by SP for which PJG has acted as a supervisor in conjunction with FS. Where possible text and figures are reproduced in

original form, however substantial reordering, alongside edits and additions have been selectively made to show how the work produced over PJG's candidature relate to one another and the wider field as well as to improve the 'narrative'. The papers are produced in full in Appendix D.

Chapter 2

Method

"Now, I return to this young fellow. And the communication I have got to make is, that he has great expectations." Charles Dickens 1861 - 'Great Expectations'

2.1 Why do we need a new modelling technique in galactic astrophysics.

In this thesis, we propose a new modelling technique further diversifying the range of semi-empirical models already used in the field. With the breadth of simulations and techniques already used, any additional models must justify their existence by showing they can succeed where others cannot. Whilst the heavy parameterisation in hydrodynamical simulation's sub-grid processes and in semi-analytic models presents a challenge to the clarity of such models, a major limitation of all the galaxy modelling techniques is the trade-off between volume and resolution. So far this limitation has severely limited the ability of models to constrain against the massive galaxy population. The reliance on discrete haloes from merger trees or N-body simulations limits the number of massive galaxies simulated in more traditional models, limiting the ability of models to compare with observations of massive galaxies found in comprehensive surveys.
STEEL has been designed as a *complementary* tool to the other galaxy models. In this chapter we describe the gaps in the current modelling space and the design choices in STEEL designed to address these gaps.

2.2 Designing to specification.

2.2.1 Volume vs. Resolution

The trade-off between volume (or the number of galaxies) and resolution (the smallest element explicitly realised in the simulation) stems directly from computational limitations. High-resolution simulations resolve orders of magnitude more small haloes than

large, below $10^{13.5}$ the halo number density increases by about one order of magnitude per decreasing decade in halo mass (left panel Figure 1.2). Due to the limited computational resources available to a simulation, there is an upper limit to the total number of haloes/galaxies that can be simulated. Either increasing the volume or lowering the resolution will increase the total number of galaxies simulated, and therefore one must come at the expense of the other. In Figure 2.1 this is visualised showing the Baryon Mass Resolution against the Number of Galaxies for many hydrodynamical simulations. The parallel diagonal lines running from top left to bottom right are lines of constant particle number which, when all other factors are constant, will correspond to computational power. The two shaded boxes show the simulations that explore two modelling regimes:

- The "zoom in", where resolution is favoured allowing for the analysis of small galaxies/haloes and the structures of larger galaxies/haloes.
- The "box", where a box of a given volume is simulated to probe the large scale structure of the Universe at the expense of resolving smaller galaxies and haloes.

Semi-Analytic and Semi-Empirical models are also similarly constrained by volume and resolution. Traditionally each uses merger trees which can either be generated analytically or extracted from an N-body simulation. Models using N-body simulations directly inherit the volume vs. resolution trade-off shown in Figure 2.1, and models using analytically generated merger trees must choose a minimum 'resolved' halo mass which will dictate the number of haloes in each tree and thus set limits on the total number of trees. However, models using merger trees have one additional level of flexibility: they can have a minimum subhalo mass and a minimum central halo mass. This provides flexibility in what can be simulated, by flexibly 'pruning' the tree and removing (sub)haloes we can adapt the halo background we are using to focus computational power on important features.

STEEL has been specifically designed to overcome the limitations of volume and mass resolutions using a *"Statistical Dark Matter Accretion history"* described in full in Section 2.3.1. In brief, the statistical accretion histories follow the average growth of a given mass of halo backwards in time from $z = 0$. At each time-step t, the average number of infalling subhaloes to a given halo growth history is calculated by comparing the USHMF at t and $t + \Delta t$. The lifetime of subhaloes controlled by the input dynamical timescales. The advantage of calculating the subhalo accretion in this way is that the prerequisites are only the average halo mass growth and the USHMF, each of which has been analytically defined [104, 105]. STEEL, therefore, simulates massive haloes and small haloes on equal footing and is able to produce massive rare galaxies in the same run as smaller galaxies with the same precision.

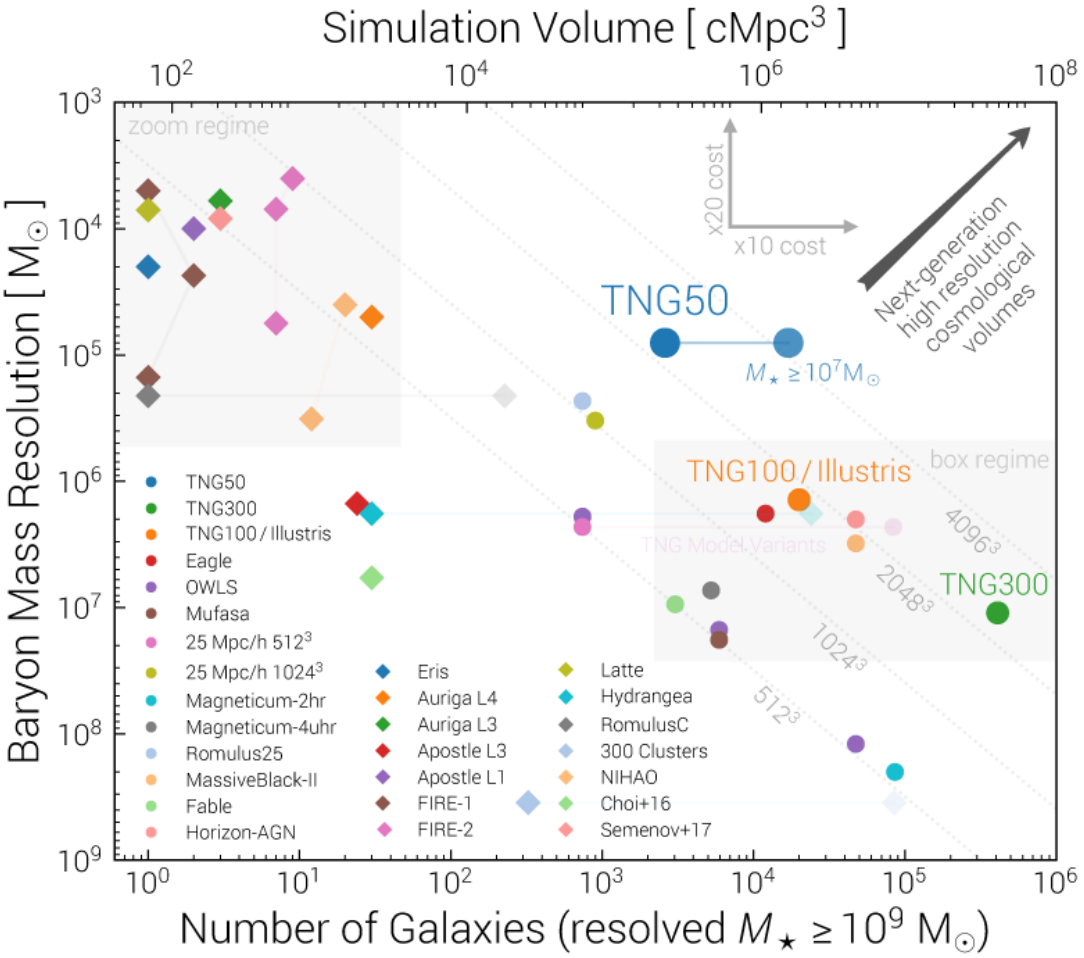


FIGURE 2.1: In this figure, the positions of several simulation is volume-resolution space is shown. The vertical axis is baryon mass resolution or the size of the elements in the simulation. The horizontal axis is the simulation volume (shown on the top) which correlates with the number of galaxies (shown on the bottom). The faint grey lines running from top left to bottom right are lines of constant particle number. The two shaded boxes in the top left and bottom right show respectively the zoom and box simulation regimes. Coloured shapes are galaxy simulations as labelled. Image used with permission from: Illustris-TNG Project, Nelson et al. [7]

2.2.2 Flexibility through stability

Hydrodynamical and Semi-Analytic models are tuned for two reasons. The first is for stability, these models have tens of parameters acting in the sub-grid for hydrodynamical or as the primary modelling tool for semi-analytic models. The issue stems from multiple parameters each contributing to one galaxy variable, changing one parameter ripples throughout the model. In abstract consider a model with two parameters, α that controls a galaxy mass with a secondary effect on the size, and β that controls galaxy colour with a secondary effect on galaxy mass. Changing β with the intent of modifying the colour of galaxies will force a reaction in α to compensate for the change in mass and then propagate into the size of galaxies. Secondly, the tuning allows fits to observations,

the models are given freedom in their parameter sets to fit many galaxy properties at the same time but due to the overlap in what each physical assumption can control this leads to degeneracy in parameter space. To complicate comparison further, some models have more physical routines than others resulting in models and parameters in the less complex model compensating for this missing physics. The diversity in models and the diversity introduced by different tuning data has been extensively discussed in the literature [106–108].

This degeneracy has in some part been addressed by semi-empirical models. Fundamental variables, such as galaxy stellar mass, are set using observationally-based techniques such as abundance matching (Section 2.4.2), which remove the reliance on the physical models and parameters that are required to generate this quantity in more comprehensive/ab-initio models. It is of note that an empirical model sets this observation and often locks the non orthogonal parameter. The degeneracy may not be physically gone simply its ability to obscure other physics in the model.

'Bottom-up' modelling like this gives semi-empirical models the ability to use dramatically reduced parameter spaces and also reproduce essential observables, such as the central stellar mass function, by design. Additional physical models are added only where necessary, creating tight constraints for the core of the model and severely limiting possible degeneracies. In the semi-empirical framework, additional physical processes and associated parameters can be added in a modular fashion. For example, Shankar et al. [34] added an analytic recipe for the size evolution of galaxies onto a network of dark matter merger trees in which central and satellite galaxies were assigned via abundance matching techniques. In the Shankar et al. [34] SEM, the addition of the "size evolution" block, is completely independent of the core features of the model, namely the mass evolution of galaxies, which is fully controlled by the dark matter assembly and the input SMHM relation. This modularity of SEMs prevents degeneracies which often affect more complex, multi-parameter models. Semi-empirical models can use this to great advantage testing, for example, multiple size evolution models on one assembly background without the need for re-tuning or fear of disrupting the essential results. In this thesis, we extend this idea of empirical stability and flexibility further. In STEEL the backbone is of statistical nature, in which the different additional physical processes are expressed in the form of probabilistic models. These models are entirely modular and their impact on the final outputs can be individually analysed.

It is noteworthy that some hydrodynamical cluster simulations are beginning to use a technique where haloes/galaxies are "genetically modified". These methods run a simulation multiple times and make slight adjustments to the initial conditions such that the models are nearly identical. For example, a change may be made to the distribution of dark matter to preference a larger merger mass ratio without changing the total simulation mass. This is a highly complementary and powerful technique that is yet to be fully

realised [109].

2.2.3 Consistency

It is well documented that reproduction of the stellar mass function at multiple epochs has been a challenge for semi-analytic models [71, 108] and hydrodynamical models (although major improvements have been seen recently, for example, Illustris vs Illustris TNG [23, 110]). This has not, however, prevented the authors of such models, particularly SAM, from making bold claims about the successes of their models in terms of morphology and mass assembly without providing a rigorously constrained stellar mass function history [e.g. 22, 32][1].

A simple analogy that highlights the issue determining galaxy assembly physics without a properly constrained stellar mass function history is as follows: A kettle modelled at temperature A is cooled under an arbitrary physical model γ, to temperature B in time T. If A is observed to be false compared to the data, and B is observed to be consistent with the data, then the inference that γ is a correct physical model is baseless. Galaxy models that fail to reproduce the stellar mass function at high redshift repeatedly invoke this kind of logic when the $z = 0$ stellar mass function is reproduced. An overabundant stellar mass function at high redshift that evolves to the observed stellar mass function at low redshift, implies either lack of growth in the galaxy population to allow the observed stellar mass function to catch up or over-merging of galaxies to reduce the total number density. A semi-empirical model designed to faithfully reproduce the time-dependent stellar mass function and the distribution of satellite galaxies as the foremost modelling priority has a major advantage.

2.2.4 Speed

Speed is a critical ingredient in any model. Hydrodynamic models have run-times of months using millions of CPU hours, and semi-analytic models constrain highly-multi dimensional parameter space for thousands of merger histories. In each case, high speed is critical to the models reaching conclusions in reasonable time frames. Semi-empirical models are faster by design. Characterised by smaller parameter spaces, they can leverage more computational time to explore physical models or take advantage of larger simulation boxes.

By using a *"Statistical Dark Matter Accretion history"* and forgoing discrete haloes/merger trees STEEL's run time is dependent on different factors to other models, notably the width of the mass bins used for the halo and subhalo mass functions (which corresponds to the resolution of haloes), and the computational complexity of any physical modelling

[1] A merger history matched to an observational merger history as provided by Hopkins et al. [32] is insufficient as we show in Chapter 5 the methods involving pair fractions used to estimate these rates have a large degree of systematic error.

routines. As STEEL is not tuning parameters in high dimensional space or applying physical models one by one to thousands of discrete merger trees, it can instead leverage computational power to test combinations of internal physical models or be run on more modest, cheaper, and widely accessible hardware.

2.3 Modules and Methods

2.3.1 Statistical Dark Matter Accretion History

Fundamentally, the single most important element of STEEL and a large part of what makes it unique, enabling an alternative view of galaxy formation problems, is the ***statistical dark matter accretion history***. As described above, a large part of the power of STEEL comes from the removal of discrete merger trees or haloes. We provide here a discussion of N-body and processed merger trees building upon Section 1.2 followed by a detailed description of how the statistical dark matter accretion history is constructed.

2.3.1.1 Traditional methods and merger trees

The state of the art in dark matter modelling is N-body simulations. N-body models simulate billions of individual particles, computing the acceleration and updating the position for each particle over thousands of time-steps. They calculate the evolution of dark matter from the distributions that mirror that of the cosmic microwave background, to the cosmic web of haloes and filaments inferred from observations of galaxies today. The non-interacting nature of **cold** dark matter allows for only gravitational force to be considered; therefore, modern dark matter simulations are able to leverage the massive power gains made recently in GPU computing allowing for ever larger and higher resolution dark matter only simulations to be run.

The outputs of even the previous generation N-body dark matter simulations, such as the Bolshoi simulation a snapshot of which is shown in Figure 2.2, total tens of gigabytes for individual 50 Mpc^3 sections and terabytes to capture the whole 250 Mpc^3 simulation. To use all of this information would be impractical for any galactic modeller without access to high-performance computing hardware and a good understanding of memory management. Furthermore, despite the excellence provided by halo-finders and merger tree extractors such as ROCKSTAR [44] and the 'consistent trees' algorithm [86], the merger trees are frequently not fit for use in models unless specifically built to handle these inputs. Finally, as discussed by van den Bosch [104, 111, 112], the interpretation of dark matter simulations is of vital importance: both the analysis tools and the fundamental design decisions taken when building the simulation, have systematic effects on the outputs all which complicate comparison between simulation outputs, e.g. subhalo statistics, velocity fields, etc.

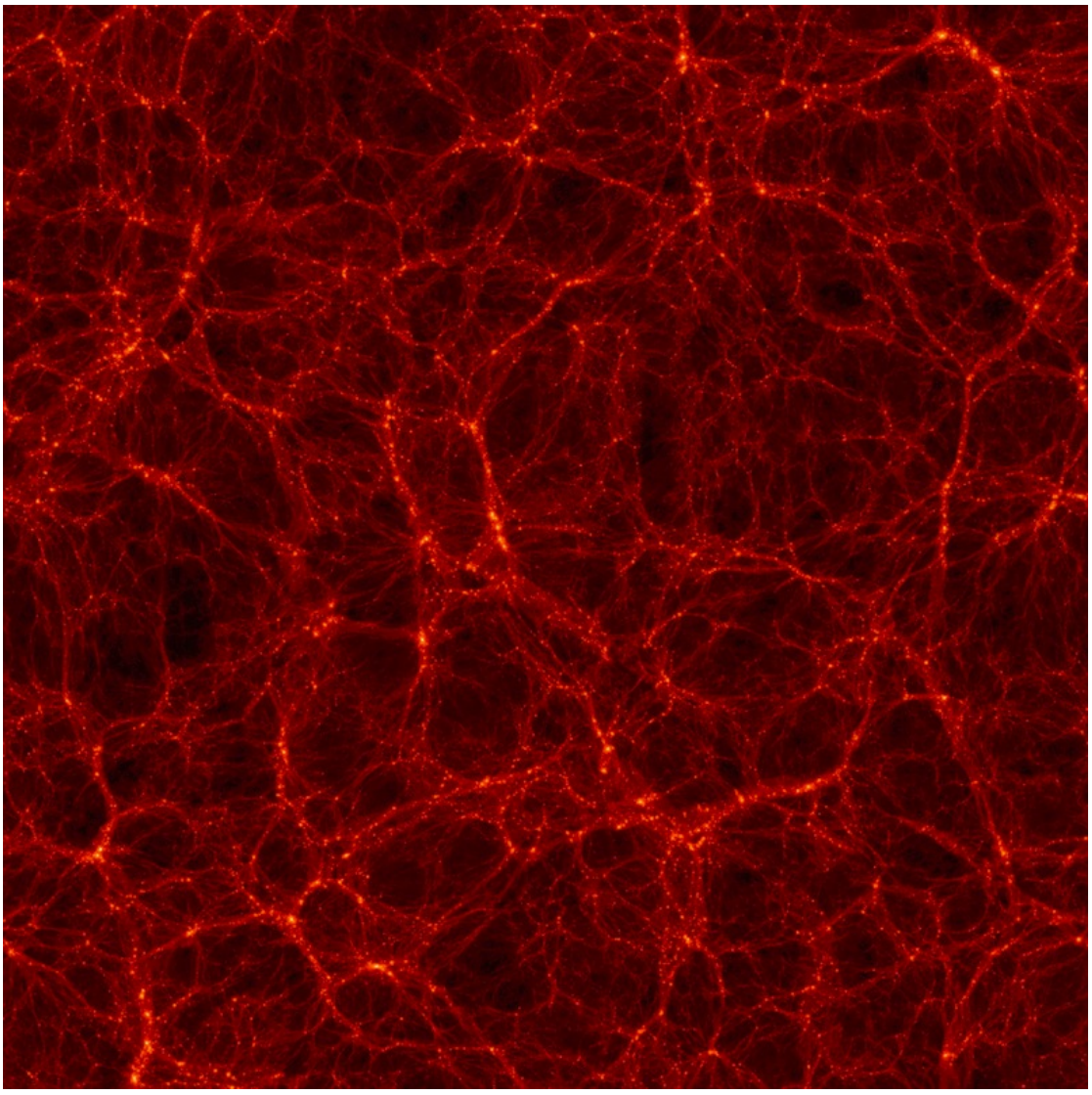


FIGURE 2.2: Snapshot of the 250 Mpc^3 box of the Bolshoi Simulation. Brighter regions indicate a higher density of dark matter. Note the clustered bright points connected by large filaments. This structure is known as the cosmic web. Image credit: Bolshoi Simulation http://hipacc.ucsc.edu/Bolshoi/Images.html

The choice of many analytic and empirical models to instead opt for analytically derived Press Schecter trees that can be generated 'on-the-fly' or pre-processed comes from the simplicity and flexibility afforded by this technique. However, trees generated in this way are still affected by volume (i.e. number simulated) and resolution (i.e. size of the smallest halo accounted for in any split) limiting there effectiveness.

2.3.1.2 State-of-the-art statistical method

The core principle of the statistical methodology is to treat parent haloes and satellites galaxies/haloes as "average" populations, avoiding the issues with volume and resolution as described previously. Furthermore, by relying on the capabilities of the halo finders

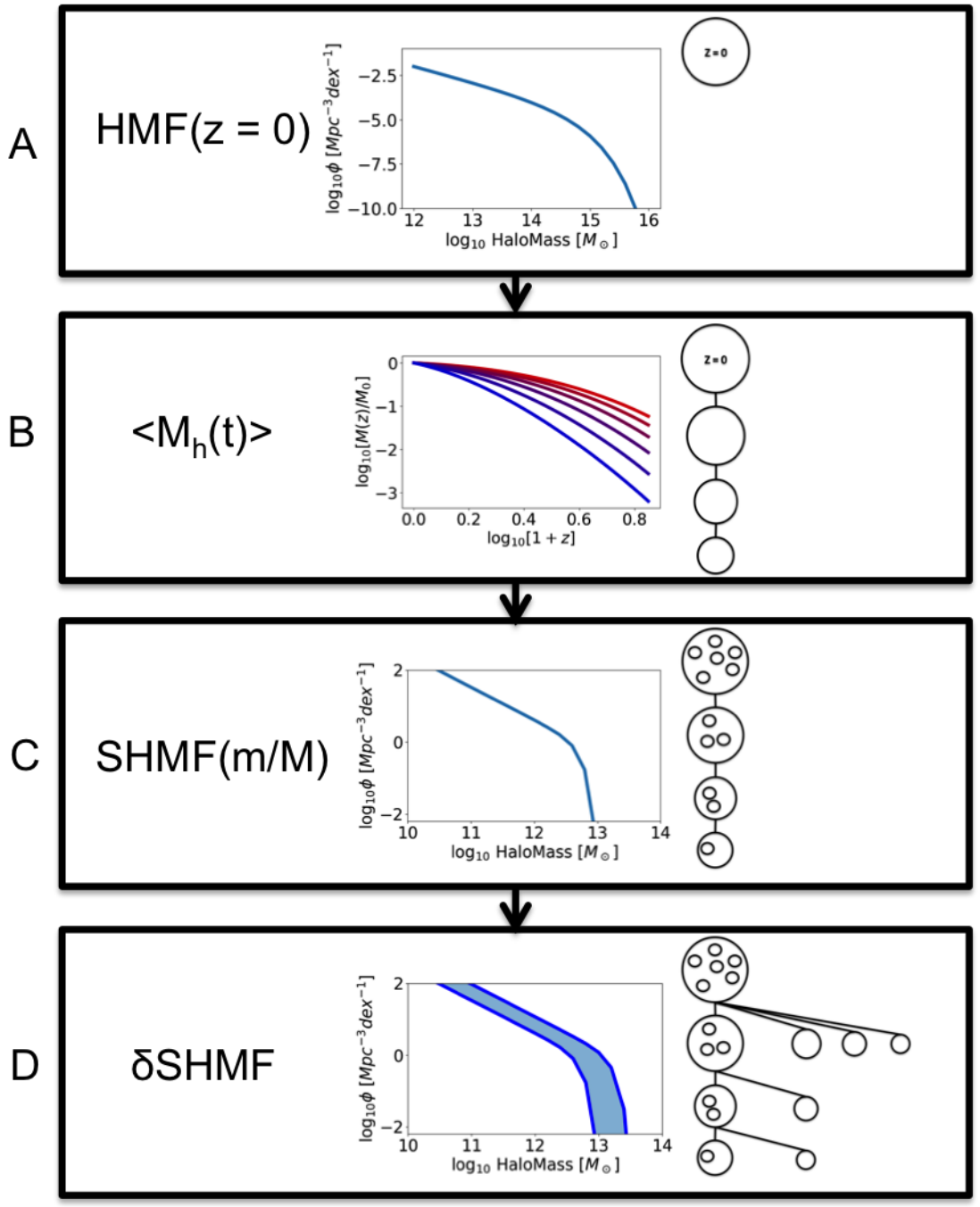


FIGURE 2.3: We show the main steps in building the statistical dark matter accretion history for STEEL. Each panel shows a feature from a traditional merger tree and the statistical function used to replace it. A: The HMF is used to calculate the number densities of central haloes. B: Average mass growth histories are used to calculate the size of each mass bin at previous epochs. C: The (unevolved) SHMF is used to populate each central at each redshift with subhaloes. D: The average number densities of accreted subhaloes at each epoch are calculated by taking the difference between each mass bin of the (unevolved) SHMF at consecutive redshift steps.

to account for the often complex histories as discussed in Section 1.2, we avoid the issues caused by such re-accreted or backsplash haloes. Here we detail the step-by-step

construction of the statistical dark matter accretion history complemented with a graphic representation in Figure 2.3.

Central Haloes

We start by considering a fine grid of central dark matter haloes ranging from $M_h = 10^{11}\,M_\odot$ to $M_h = 10^{15}\,M_\odot$ at redshift $z = 0$. Their number densities are given by the halo mass function (HMF) of Despali et al. [113], which is obtained using the COLOSSUS Python package [114]. COLOSSUS also contains spherical over-density conversions required throughout this work, as well as many other halo mass functions which may be switched to or from to test different cosmologies as the statistical dark matter accretion history is theoretically independent of may of the choices.

The halo mass function provides the number densities of haloes in a given mass bin (Figure 2.3, Panel A). The average mass growth histories of all main progenitors in the bin of halo mass $[M_h, M_h + dM_h]$ are then calculated using the analytic model from van den Bosch et al. [104][2]. This provides the average "main progenitor" branch of a traditional merger tree for each mass bin at $z = 0$ (Figure 2.3, Panel B.)

Assigning Subhaloes to Parent Haloes

In order to predict the number of satellite galaxies, we must associate to each parent/-central halo the number and mass of subhaloes they are expected to contain. To achieve this we use the subhalo mass function (SHMF). The SHMF describes the expected distribution of subhaloes of mass $M_{h,sat}$ in a given parent halo of mass $M_{h,cent}$, as a function of $M_{h,sat}/M_{h,cent}$. An average treatment of subhaloes is supported as it has been found that over a range of host masses [115] and cosmologies [105] the subhalo pospulations are simmilar. Multiple definitions for the SHMF exist depending on the way a subhalo is defined. In this work, we use two definitions of the SHMF. The first is the unevolved SHMF (USHMF), which describes the total subhaloes accreted over a parent halo's lifetime. In the unevolved SHMF any merging or stripping in the subhaloes occurring after infall is ignored. Several groups have been able to constrain the unevolved SHMF using the infall masses of subhaloes in N-body simulations [105]. In what follows, we use a recent rendition of the unevolved SHMF by Jiang and van den Bosch [105], which is calibrated against the Bolshoi simulation[3].

The second definition we use in this work is the unevolved "surviving" SHMF (USSHMF). Subhalo masses are assumed "frozen" at infall but the subhalo number densities can reduce compared to the unevolved SHMF as the unevolved surviving SHMF accounts for subhalo disappearance due to merging with the parent halo. Subhalo merging

[2]This model further improves on the seminal work by Parkinson et al. [51], which was aimed at reproducing numerical merger trees, optimised with small redshift steps minimising the development of systematic errors at late cosmic epochs.

[3]We direct the interested reader to Jiang and van den Bosch [105] for further discussion of the unevolved SHMF as well as of other SHMFs, such as the evolved SHMF where the number densities are affected by both subhalo stripping and mergers.

is due to loss of angular momentum to the parent halo via 'Dynamical Friction' described in Section 2.3.1.3, and the analytic details of subhalo dynamical friction are given in Section 3.2.1. We show in Figure 2.4 for a representative parent halo of mass $\log M_{h,cent}/M_\odot = 12.80$, the unevolved SHMF and three unevolved surviving SHMF characterised by different dynamical friction timescales τ_{dyn}. Larger τ_{dyn} lead to a milder reduction in subhalo number densities as subhaloes take longer to merge with the parent halo. Lower τ_{dyn} are less effective in reducing the number densities of smaller subhaloes which are more likely to have dynamical friction timescales comparable to or larger than the Hubble time at $z = 0$.

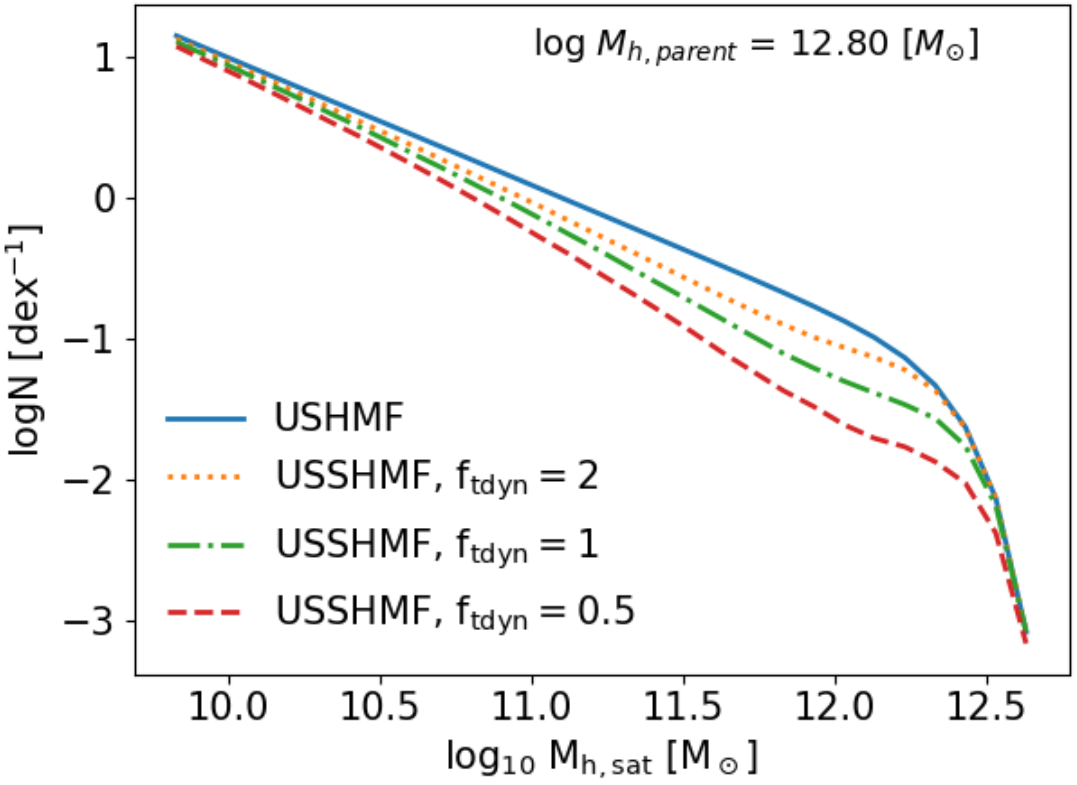


FIGURE 2.4: Comparison between the unevolved SHMF (solid line) and three unevolved surviving SHMF (dotted lines) for a parent halo of mass $\log M_{h,parent}\ [M_\odot] = 12.80$. The factor $f_{t_{dyn}}$ is applied to the merging timescales of the haloes. Lower factors correspond to lower unevolved surviving SHMF where more subhaloes have merged.

Average Subhalo Accretion

At each redshift step along with the mass growth histories, we calculate the unevolved SHMF associated to the parent halo mass. This is equivalent to the substructure found in traditional merger trees (Figure 2.3, Panel C). However, unlike traditional methods, our statistical approach is able to probe "rare" subhaloes without running prohibitively large volumes of merger trees.

For each time step we can now calculate a mass function describing the number density of subhaloes accreted onto the population of central haloes in the halo mass bin $[M_{h,cent}(z)$,

$M_{h,cent}(z) + dM_{h,cent}(z)]$, which is computed by differentiating the unevolved SHMF across two neighbouring redshift steps z and $z + dz$. In practice we can calculate the average number density of subhaloes of any given mass $M_{h,sat}$ that are accreted in the redshift interval dz onto the main progenitor haloes with mass in the bin $[M_{h,cent}(z)$, $M_{h,cent}(z) + dM_{h,cent}(z)]$ as,

$$\begin{aligned} &\delta USHMF[z, M_{h,cent}, M_{h,sat}] = \\ &USHMF\Big(\frac{M_{h,sat}}{M_{h,cent}(z)}\Big) - USHMF\Big(\frac{M_{h,sat}}{M_{h,cent}(z + \delta z)}\Big). \end{aligned} \tag{2.1}$$

In this way the unevolved subhalo accretion history ($\delta USHMF$) is retrieved for all main progenitor haloes at all redshifts.

2.3.1.3 Dynamical Friction

Dynamical Friction is the process by which an object (particle) moving through a mass 'field' (or distribution of particles) loses velocity and increases the velocity of the field via gravitational interactions. A simplified example is shown in Figure 2.5; the red particle is initially moving to the left between two rows of particles of an infinite extent that interact only through gravity (and only with the red particle). The particle in Panel A receives an equal pull from all particles and therefore experiences no acceleration. The black particles directly above and below the red are the most strongly attracted and therefore move the greatest distance toward the red, with particles that are further away moving less proportionally to their distance. By Panel B the red particle has moved to the left and due to the previous effects now has more mass clustered it along the path. The break in the symmetry exerts a net force on the red particle in the opposite direction to its motion. In Panel C, where more black particles have been drawn behind the red particle the effect is magnified and the resultant force strongly opposes the motion.
The thought experiment described above captures a number of the key features characterising the dynamical friction process:

- The red particle experiences a 'drag' force induced by the particles it draws in behind it, eventually losing most of its momentum.
- The particles are non-interacting aside from gravity, in line with the leading dark matter theories.
- The energy of the red particle is dissipated into the velocities of the black particles: The motion of a galaxy through a dark matter field increases the random motion of the dark matter.

The main features missing from the thought experiment are:

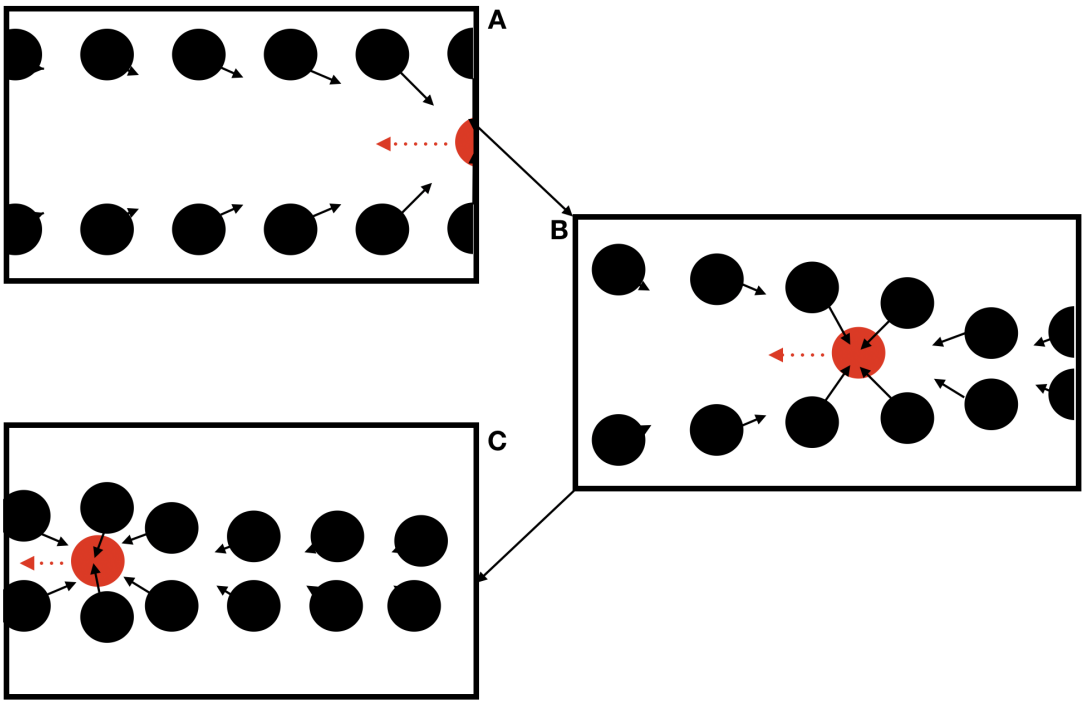


FIGURE 2.5: In this handmade cartoon, we depict a thought experiment to demonstrate the process of dynamical friction. We show a red test particle moving through two rows of black particles that are infinite in extent. The particles interact only through an attractive force that reduces with distance akin to gravity. The black particles interact only with the red particle and not with each other. The black particles start at rest and the red particle starts with some arbitrary velocity to the left. Panel A shows the starting configuration of the particles, the panels B and C show two theorised later states of the system. For an animation of a simulation designed around this thought experiment see: https://twitter.com/RSEGrylls/status/1229104848643198979 (made and shared by PJG)

- The dark matter field is not uniform. It is a potential well with more mass in the centre.
- The red 'test' particle is in-fact a distribution of masses that are gradually lost to the field.
- The red particle moves in an orbit around the centre of the potential well and is stripped of mass, velocity, and momentum.

2.4 Utilised Data

In this thesis we use the stellar mass functions defined below, along with the halo mass functions, to constrain the SMHM relationship. In addition to the SMF, at high redshifts, we use measurements of massive clusters to constrain the performance of the model to reproduce the satellite galaxy distributions.

2.4.1 Stellar Mass Functions

2.4.1.1 Low Redshift, $z = 0.1$

At low redshift we use the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7) from Meert et al. [116]. The data from the SDSS-DR7 spectroscopic sample [117] contain $\sim 670,000$ galaxies fitted with a Sérsic + exponential model [PyMorph; 116] with associated halo masses and central satellite classifications from [118]. The improved photometric analysis by Meert et al. [116] provides more reliable estimates of the stellar mass function at the high mass end which are more abundant than previous estimates [119, 120]. In this thesis, the effect of the enhanced high mass end on galaxy assembly is investigated. We compare to previous determinations of the stellar mass function using as an example the de Vaucoulers [121] based cmodel fits from SDSS [117]. The latter definition of galaxy stellar mass has been extensively discussed not to be accurate, partially due to incorrect sky subtraction and adoption of non-ideal light profiles [122]. Bernardi et al. [120] have clearly shown that the choice of light profile is not a simple matter of "semantics". The single or double Sèrsic models perform better in fitting the surface brightness of galaxies independently of the galactic environment [116]. The performance is thus not related to the inclusion of the intra-group or intra-cluster light in the fit [120].

2.4.1.2 High Redshift, $z > 0.1$

At higher redshift (0.3 <z <3.3) we use stellar mass functions from the COSMOS2015 catalogue [123]. Here masses are defined using spectral energy distribution fitting, including ultra-deep infrared photometry. Davidzon et al. [123] use Bruzual and Charlot [124] stellar population synthesis models to estimate stellar masses. As SED fitting is notably different from light profile fitting, one cannot apply the same corrections as in Mendel et al. [125]. Nevertheless, to match the mass-to-light ratios adopted by Mendel et al. [125], based on the Bell et al. [126] mass-to-light ratios, we follow Bernardi et al. [122] and increase the Davidzon et al. [123] stellar masses, based on Bruzual and Charlot [124], by +0.15 dex. We note that the resulting $z = 0.37$ stellar mass function after this correction is in remarkably good agreement with the $z = 0.1$ stellar mass function by [122]. Our result also matches the findings by Bernardi et al. [119], who showed that, by making use of the BOSS sample, the stellar mass function shows negligible number density evolution up to $z \sim 0.5$.

2.4.1.3 Cluster at z = 2.5, Wang+ 2016

The highest redshift cluster we compare to is a $M_{200c} = 10^{13.7} M_{\odot}$ halo containing 15 galaxies with $M_* > 10^{10} M_{\odot}$ at a redshift of $z = 2.5$. This cluster is reported in Wang et al. [12], and we provide a brief description of the observation and data here. The

cluster is observed using IRAM-NOEMA, VLT-KMOS, VLA, XMM-Newton and Chandra for the spectroscopic observation and redshift determination. The galaxy masses are determined assuming a Salpeter [127] IMF, which we correct to a Chabrier [128] IMF, by decreasing the stellar masses by 0.24 dex. The halo mass ($M_{vir} \sim 10^{13.93} M_{\odot}$) of the cluster is estimated in three different ways, using the total X-ray luminosity, the velocity dispersion of its member galaxies above $M_* = 10^{10.76} M_{\odot}$, and the stellar richness of the cluster [4]. Given this object was a targeted cluster, we cannot estimate the cosmic abundance (i.e, the number per cubic megaparsec). For analysis and comparison later in this work, we assign this cluster an abundance of $N(> M_* = 10^{13.93}) = 10^{-7.15}$ $[Mpc^{-3}]$ which is estimated by integrating the halo mass function in the limits $[10^{13.93}, \infty]$, thus providing an upper limit to the number densities associated to clusters of this mass.

2.4.1.4 1959 Clusters at z = 0.7 - 1.0, Wen & Han 2018

We compare to the cluster sample from Wen and Han [13], which contains 1959 clusters from SDSS-DR14 [129] and the WISE survey [130]. The clusters are identified in the W1 band, and foreground objects are removed using the SDSS photometric data. The cluster mass and richness are estimated using the total W1 band luminosity within 1 Mpc of the central galaxy. As performed above, to each cluster we assign an upper limit to their abundances from the cumulative integration of the halo mass function.

2.4.2 Abundance matching

In this work, we populate dark matter haloes with galaxies using the abundance matching technique where galaxies are assigned to haloes by comparing the relative abundances of galaxies and haloes. For example in Figure 2.6 the horizontal lines connect points of constant number density between the HMF (red, top left) and the SMF (green, bottom right). Halo/stellar masses with corresponding number densities are then used to define a mapping between the masses. This connection is called the stellar-mass-halo-mass (SMHM) relationship (black, bottom left).

The abundance matching used in STEEL extracts central haloes from the halo mass function from Despali et al. [113] obtained using COLOSSUS[114], and a subhalo mass function subdivided by the redshift of infall generated from the statistical dark matter accretion history used by STEEL. Subhaloes are assumed to follow the central SMHM relation at infall. We simplify our abundance matching by using a frozen model such that baryonic evolution after infall (stripping, star formation, etc.) is not included. The latter assumption provides a good approximation as after infall the dominant factor determining the

[4] We note the velocity dispersions and X-ray luminosity estimations give the cluster mass as $M_{vir} = 10^{13.73} M_{\odot}$ and the estimate given by mass richness is significantly higher $M_{vir} = 10^{14.6} M_{\odot}$. Whilst we used the published average the lower cluster mass is in as-good or better agreement with model results.

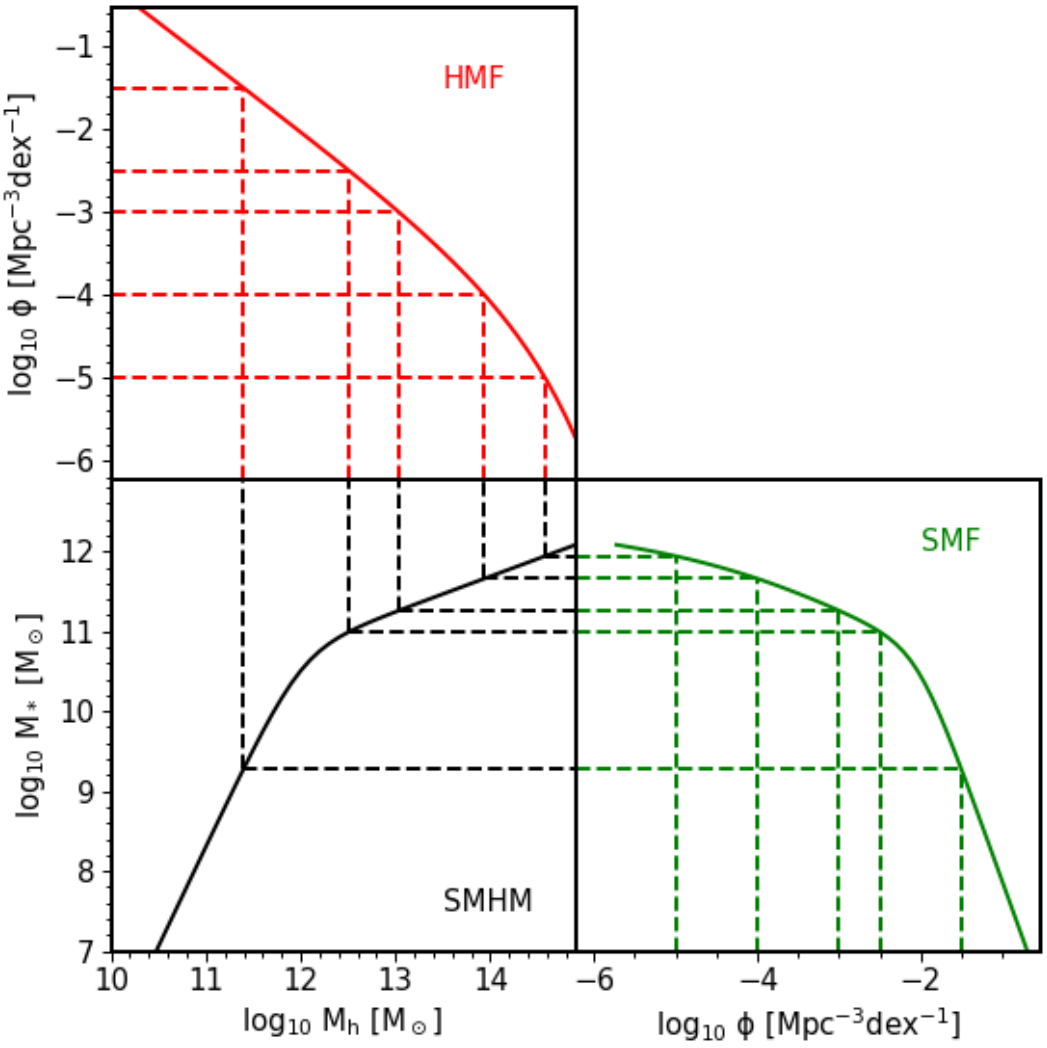


FIGURE 2.6: A cartoon to show matching the HMF (top left) and the SMF (bottom right) via abundance; the mapping between stellar mass and halo mass, referred to as SMHM relation, (bottom left) is created.

abundances of satellite galaxies is the dynamical time and not evolutionary processes (Paper I).

To fit stellar mass functions over multiple epochs we convolve our halo mass functions with a parametric SMHM relation similar to that proposed by Moster et al. [131],

$$\begin{aligned} M_*(M_h, z) &= 2M_h N(z)\Big[\Big(\frac{M_h}{M_n(z)}\Big)^{-\beta(z)} + \Big(\frac{M_h}{M_n(z)}\Big)^{\gamma(z)}\Big]^{-1} \\ N(z) &= N_{0.1} + N_z\Big(\frac{z-0.1}{z+1}\Big) \\ M_n(z) &= M_{n,0.1} + M_{n,z}\Big(\frac{z-0.1}{z+1}\Big) \\ \beta(z) &= \beta_{0.1} + \beta_z\Big(\frac{z-0.1}{z+1}\Big) \\ \gamma(z) &= \gamma_{0.1} + \gamma_z\Big(\frac{z-0.1}{z+1}\Big). \end{aligned} \tag{2.2}$$

In what follows we create SMHM relations using both the cmodel and PyMorph SMF described in Section 2.4.1 at redshift $z = 0$ to constrain the parameters N, M, β, and γ (normalization, knee, low mass slope, and high mass slope). We use only the central stellar mass function, using the [118] central/satellite identification, and central halo mass function. The fit is performed using a Markov Chain Monte Carlo (MCMC)[5], implemented using the PYTHON package EMCEE [132], over a large parameter space ($P_{M,N,\beta,\gamma}$) covering all four parameters. Given a point in parameter space $P_{M_i,N_i,\beta_i,\gamma_i}$, the stellar mass function is constructed using the halo mass function and the SMHM relation. Each bin of the central halo mass function is associated with a Gaussian distribution of stellar mass via the SMHM relation with scatter 0.15 dex. This distribution is multiplied by the halo mass number density to convert to galaxy number density which is added to the relevant stellar mass bins of the stellar mass function in construction. This operation is then repeated overall mass bins of the halo mass function to produce the complete central stellar mass function. For each point, $P_{M_i,N_i,\beta_i,\gamma_i}$ in the parameter space, the stellar mass function associated to that point is compared via a likelihood function to the observed stellar mass function to provide the MCMC with the probability that the given point is the 'true' SMHM relationship.

In Figure 2.7 the above process is visualised. Stripes in halo mass function are selected as shown in the left-hand plot and then again in the middle as red bands. By propagating these bins of halo mass through the SMHM relation and associated scatter they each create a Gaussian of stellar mass as shown by overlapping green shading in the middle panel. Once weighted by the median halo number density they are added as green solid

[5] In actuality, the fit is first done by hand correcting parameters one at a time and visualised at each step. In doing so an appreciation of the connection between parameter value, SMHM shape, and SMF is built that promotes an in-depth understanding of the system and its caveats. Around the minimum value, a brute force search is performed with various ranges and resolutions. This provides a test of the fit estimator that will be included in the MCMC. Once the limitations of the method are known, building the MCMC is more robust as sensible checks and balances are included. The MCMC will often provide a solution close to the one found by the previous methods but we also include the robustness/errors associated with the result.

lines in the rightmost panel. Summing up each of the Gaussians an estimation of the stellar mass is achieved. This animated figure is available at https://twitter.com/RSEGrylls/status/1231279549704474624. From this figure we see several features of the SMHM relation:

- From the visualisation, it can be seen that the scatter in the SMHM relation has two major effects: Firstly, it can smooth out jaggedness introduced by the binning. One should, therefore, ensure that the scatter parameter is not influenced by the size of the binning to confirm results are physical and not numerical artefacts. Secondly, the scatter dictates the size of the stellar mass Gaussian elements, it thus ultimately controls the range of stellar masses within a given halo mass bin. The width of the scatter has a second effect whereby lower mass halo bins can contribute to higher mass stellar mass bins, a process routinely known as 'upscatter'. Due to the greater number densities of lower mass haloes, the upscatter effect can greatly impact the number densities of massive galaxies. This can be seen at high stellar masses where the peak of the stellar mass distribution sits below the previous distribution. This is the reason the scatter and the high mass slope are degenerate parameters when creating massive galaxies [133].

- It can be seen from the animated figure, or by tracing haloes of mass $10^{12} M_{\odot}$ through the middle panel, that the knee of the SMHM relation is closely related to the high mass cutoff in the SMF.

- Due to the fact that the high-mass slope of the SMHM relation is much shallower than the low-mass slope, stellar mass distributions created are far more clustered, and thus the high mass slope controls a much smaller range of galaxies than the low mass slope.

We then fit to the Davidzon et al. [123] data both uncorrected and corrected for the cmodel and PyMorph fits respectively (see Section 2.4.1 for details). At high redshift we use the central and subhalo mass functions initialising satellites at infall as described above[6]. For central haloes, the method is the same as detailed above. However, as we use the total stellar mass functions at high redshift, we also include the total unevolved surviving subhalo mass function in the abundance matching to account for the satellite galaxies present in the observed SMF. We assume that a halo before infall hosts a central galaxy; under this assumption we use the central SMHM relation to assign satellite galaxy stellar mass at the point of accretion. For the latter, we must have information about the redshift of infall for subhaloes. We obtain from STEEL the unevolved surviving subhalo mass function as contributed by each redshift of infall. Each contributing part

[6]Ideally, as for low redshift, we would use the centrals only as we are primarily concerned with the central SMHM relation. However, lacking a well-defined central stellar mass function at high redshift, this method represents a reliable way to extend the model to higher redshifts.

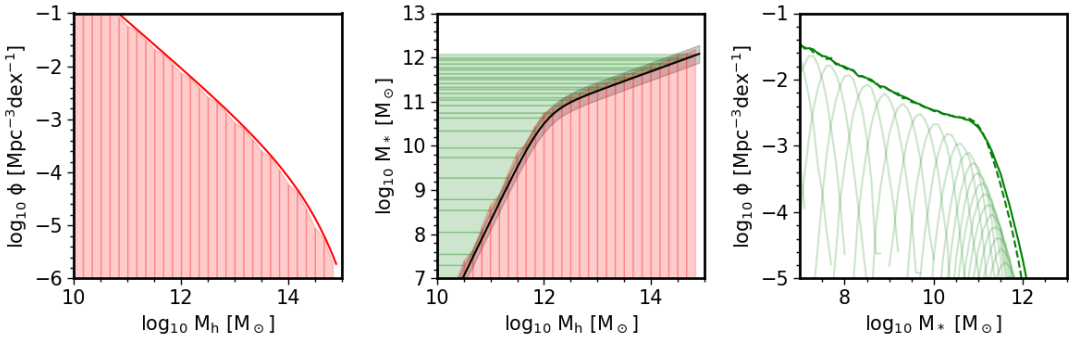


FIGURE 2.7: A cartoon to show how the 'statistical' approach can be taken to transform a HMF to a SMF via number density propagation. From left to right: Halo mass function with selected binning shown by shaded bands; The SMHM relation with the halo mass bins shown as red bands and the associated stellar mass distributions shown as green bands; The SMF with the Gaussian distributions from each halo mass bin shown as solid green curves that are then summed to create the total stellar mass function. An animated version of this plot made public by PJG can be found at https://twitter.com/RSEGrylls/status/1231279549704474624.

	M_n	N	β	γ
$\text{cmodel}_{z=0.1}$	$11.91^{+0.40}_{-0.34}$	$0.029^{+0.018}_{-0.013}$	$2.09^{+1.21}_{-1.02}$	$0.64^{+0.11}_{-0.10}$
cmodel_{zevo}	$0.52^{+0.24}_{-0.19}$	$-0.018^{+0.005}_{-0.004}$	$-1.03^{+0.049}_{-0.34}$	$0.084^{+0.20}_{-0.14}$
$\text{PyMorph}_{z=0.1}$	$11.92^{+0.39}_{-0.36}$	$0.032^{+0.016}_{-0.012}$	$1.64^{+0.85}_{-0.73}$	$0.53^{+0.11}_{-0.11}$
PyMorph_{zevo}	$0.58^{+0.15}_{0.19}$	$-0.014^{+0.007}_{-0.006}$	$-0.69^{+0.29}_{-0.36}$	$0.03^{+0.154}_{-0.147}$

TABLE 2.1: The abundance matching results for the cmodel and PyMorph data. The errors are the 16th and 86th percentile from the MCMC fiting.

is calculated using the SMHM relation at the redshift of infall and added to the central stellar mass function using the same method as with the centrals. The total stellar mass function is compared, at each redshift step available, to the data via the likelihood function to give the probability that the given point is the 'true' evolution parameters. The abundance matching best-fit parameters and associated errors for both the cmodel and PyMorph are given in Table 2.1, and plots showing the cross-sections of the parameter space are shown in Appendix C.

In Figure 2.8 we show the results of our abundance matching to the PyMorph and cmodel central stellar mass functions. The PyMorph fit is steeper above the knee compared to either the cmodel or the Moster et al. [8] model fits, as expected given the larger number density of massive galaxies found applying the Sérsic-Exponential model [eg., 34, 134]. The low mass slope for both PyMorph and cmodel are almost identical as the galaxies in this range are not affected by the photometric choice. Differences between the fits from this work and Moster et al. [8] are due to our selection of using only central

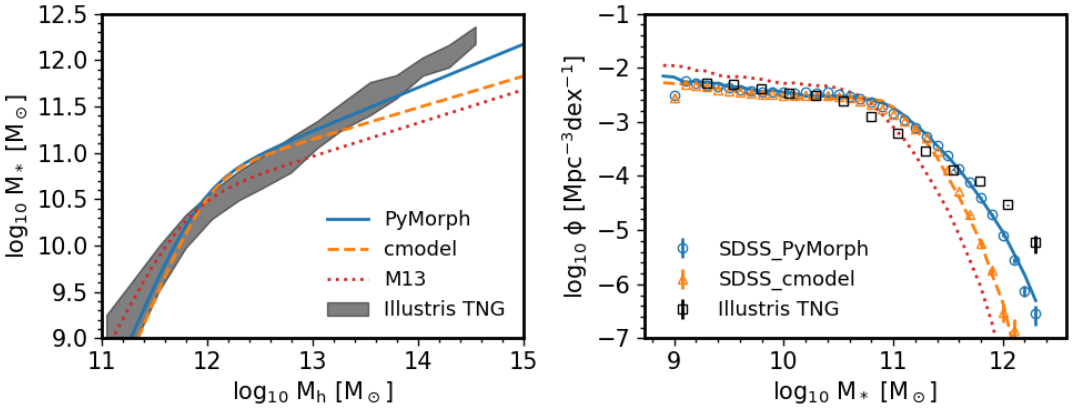


FIGURE 2.8: Left: The SMHM relation at redshift $z = 0.1$. The PyMorph (blue solid line) and cmodel (orange dashed line) fits from this work are both for central haloes/-galaxies, the fit from Moster et al. [8] (M13, red dotted line) is for all haloes/galaxies. The grey band is the relation from Illustris TNG100. Right: Stellar mass functions created using the central halo mass function and the three SMHM relations compared to PyMorph (blue circles) and cmodel (orange triangles) central stellar mass functions. The black squares are the stellar mass function from Illustris TNG100.

haloes/galaxies as opposed to the total population, and their stellar mass functions shown in the right-hand panel, are therefore missing massive galaxies.

2.4.3 Continuity star formation rate

A substantial modelling problem arises from the continuity between the evolution of the SMF and the observed SFR. The integral of the SFR over time, after correcting for stellar mass loss, should be equal to the growth in the stellar mass function. However, if the observed UV SFR is used to evolve high redshift SMF to lower redshift the result is a local SMF which far exceeds what is independently observed. This discrepancy between integrated SFR and observed SMF naturally implies that any model tuned to fit one of the two observables will necessarily fail to match the other one.

The continuity star formation rate is a theoretical quantity calculated such that the observed growth of the stellar mass function is conserved. As illustrated in Figure 2.9 by connecting two SMF at equal number densities at times 't' and '$t + \Delta t$', the growth in mass ΔM over the time Δt is obtained. In the simplest approximation, the SFR is equal to $\Delta M/\Delta t$. However, the observed star formation rate is higher than this value. As new stars are created a proportion of the mass is recycled through supernova 'quickly' in terms of cosmological time. This mass recycling given in Moster et al. [17] is used to amend the star formation rate, the rate of stellar mass loss is calculate from FSPS package from Conroy et al. [135, 136]. To account for the mass loss based on the entire star formation history of the galaxy. The fraction of mass lost by a population of stars over time τ_{ml} is given by,

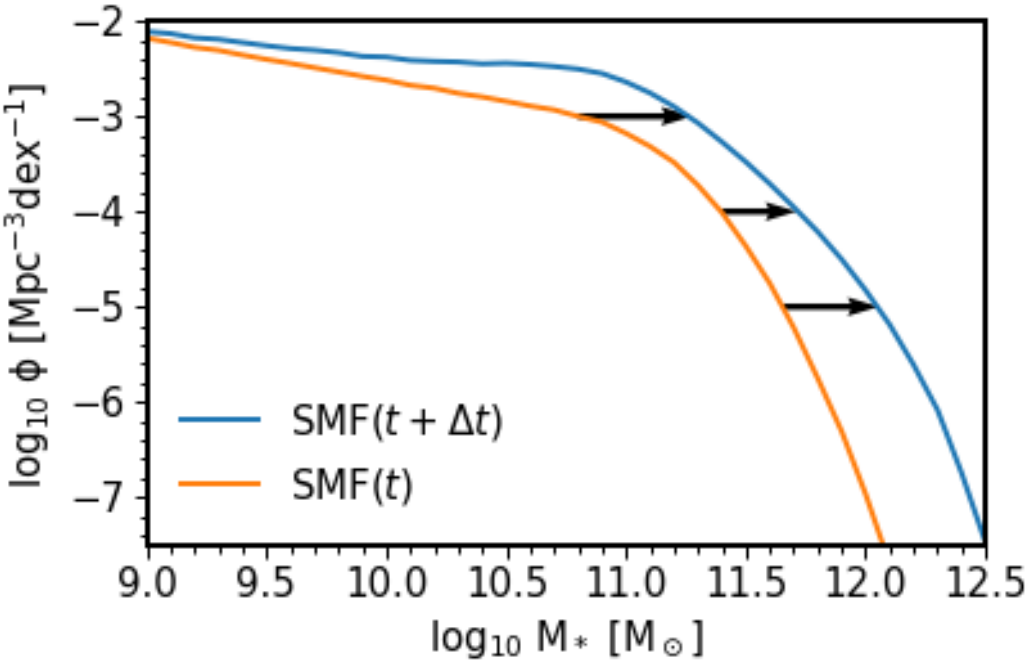


FIGURE 2.9: The continuity approach connects at constant number density galaxy populations across cosmic time. Between two epochs 't' and '$t + \Delta t$' the SMF grows differently at different number density. The mass difference, ΔM, indicated at three number densities by black arrows is the expected mass growth for each population.

$$f(\tau_{ml}) = 0.05 \ln\left(\frac{\tau_{ml}}{1.4Myr} + 1\right). \tag{2.3}$$

Summing the difference in fraction lost in a time step δt for every star formation epoch in the galaxies history (SFH) gives the mass loss rate (MLR),

$$MLR(t) = \frac{\sum_{t'=t_{inf}}^{t} SFH(t')(f[t' - (t - \delta t)] - f[t' - t])}{\delta t}. \tag{2.4}$$

The continuity SFR is therefore calculated to be $\Delta M/\Delta t$ + MLR. The final correction made to this star-formation rate is that galaxies also accrete mass via mergers. This accreted mass term (M_{acc}) is prominent for massive galaxies $M_* > 10^{11} M_\odot$. The continuity SFR at time t' is then finally given by,

$$SFR_{continuity}(t') = \frac{\Delta M(t')}{\Delta t} + MLR(t') - M_{acc}(t'). \tag{2.5}$$

2.4.4 Empirical central galaxy growth

Using the stellar-mass-halo-mass (SMHM) relation and halo mass growth histories (HMGH) we make an empirical (observationally-driven) prediction of the average central galaxy growth. In Figure 2.10 we show in the top left the HMGH for two haloes $M_{z=0} =$

$10^{14}, 10^{12.5} M_{\odot}$, the middle panel shows the SMHM, where the shading shows the extent of the evolution of the median of the relation over the redshift range. Finally, the bottom left shows the stellar mass growth history (SMGH) predicted by the latter two inputs.

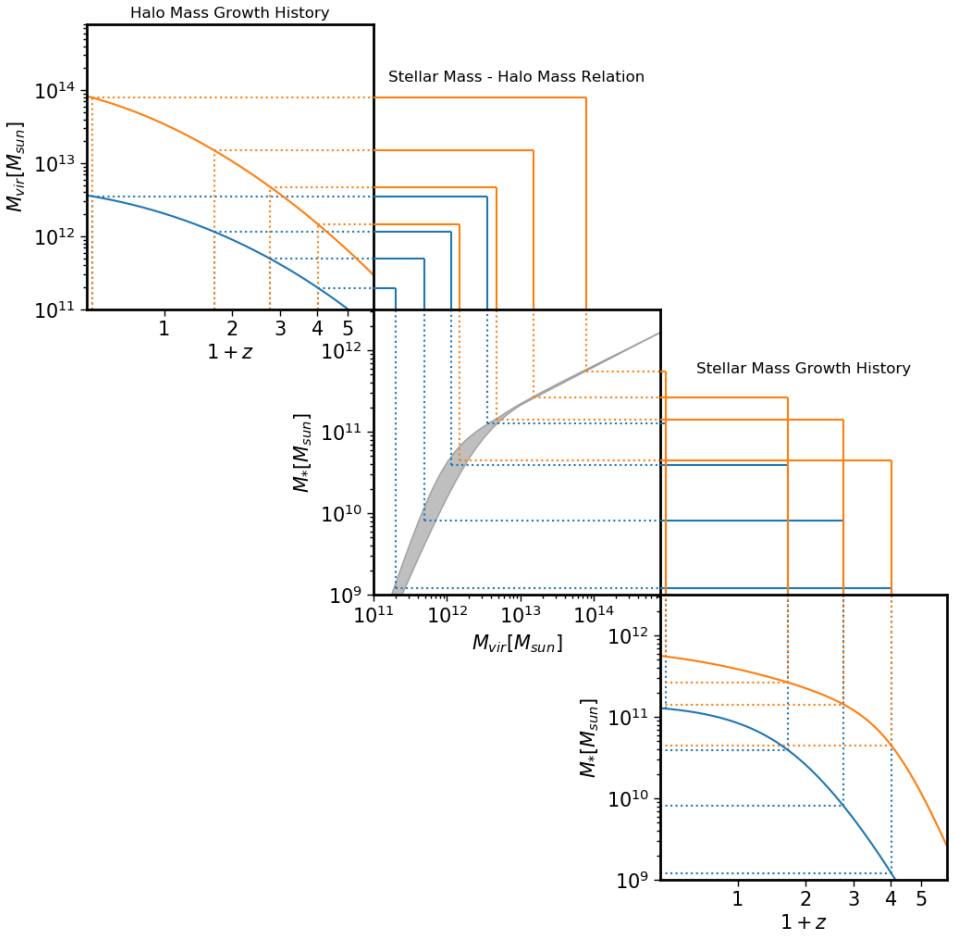


FIGURE 2.10: The Halo Mass Growth Histories (HMGH, top-left) are propagated through the redshift dependent stellar-mass-halo-mass relation (SMHM, middle) to produce corresponding Stellar Mass Growth Histories (SMGH, bottom-right). The lines illustrate matching points in redshift and the intersection with the SMHM relationship. The width of the SMHM relationship shows the extent of evolution with time.

At several redshift epochs, the HMGH are connected to the SMGH following the colour-coded lines. It is found the shape of the SMHM relationship is the primary factor that dictates the shape of the SMGH.

By identifying for each growth history the redshift at which the central galaxy stellar mass growth rapidly slows, $z = 3$ for orange and $z = 1$ for blue. We can trace the stellar mass at this point to where they are in their HMGH and to their location on the SMHM. For bothe the mass histories shown here, the HMGH is still increasing and the reduction

of stellar mass growth occurs just before the 'knee' of the SMHM relation. This change in the growth behaviour is thought to be an effect of active galactic nuclei suppressing the formation of stars in a galaxy [e.g. 137, 138]. However, as the knee is a function of halo mass, one could also attribute this to the halo mass at which infalling gas is shock heated and can no longer efficiently accrete to the central galaxy [e.g. 137, 139].

2.5 Summary and Future

This chapter has outlined the core methodology of STEEL. For exact details of the development of the methods, we refer the reader to the full papers in Appendix D. For example, the continuity star formation rate in Paper I was derived by generating SMF at subsequent time-steps and then following populations at consistent number density. In Paper II the method was updated to instead follow the central dark matter growth histories.

At its inception STEEL was a short test script used to test what the maximum number of satellites in a halo could be using given a SMHM relation. The 75 line code in Section 2.6 is a recreation of the original script with additional formatting and commenting[7].

2.5.1 Future Scientific Applications

STEEL, as it stands, has made exceptional contributions to galaxy assembly. Properly developed as detailed in Section 2.5.2 it could unlock additional capabilities that would open more scientific applications. For example:

- STEEL accurately captures the number densities of satellite galaxies in groups and clusters, as shown in Chapter 3; a natural extension to this is assigning probabilistic spatial information to these mock distributions to create estimations of traditional clustering estimators such as the two-point correlation function.
- Including additional empirical routines, e.g. the $M_{BH} - \sigma$ relation, is both simple thanks to the flexible design and would quickly produce further scientific results.
- Fitting new SMF will create different stellar-mass-halo-mass relationships that can substantially alter results as can be seen in Chapters 4 & 5.

2.5.2 Future Development

As will be seen in the following Chapters (3, 4, & 5), the novelty of STEEL provides a view on galaxy formation physics that can refresh and re-frame some ideas in the field.

[7]Note this script will not run as the SMHM relation called from SEM is not present neither are the parameters to be sent to that script. The naming conventions are as originally used being a patchwork of other scripts. trapz is a simple trapezium integrator that would also need to be replaced. This script is unlicensed and only intended as a guide to the original thoughts that went into building the foundation of STEEL

However, as STEEL grew naturally from the proof of concept, it has now reached a stage common to many large pieces of code where it requires a critical tear-down and refactor. Developing STEEL efficiently is possible only for PJG and those trained by PJG without a significant time investment, therefore a refactor to provide the following usability and science benefits should be considered a priority.

- Development ease: A redesign of the software architecture will make the addition of new modules easier important for retaining flexibility.
- Method: Many of the routines in STEEL involve weighted statistical distributions (mostly narrow Gaussians) it would, therefore, be in the interests of the model to redesign the methodology around an object-orientated approach that handles many probability distribution convolutions by design.
- Speed: STEEL is much faster than most models, however, the choice of PYTHON as the language has had certain drawbacks. Iterative processes are slow, for this reason, the entire star formation module was written in CYTHON a type defined version of PYTHON that is pre-compiled into C. This process is both slow and difficult. With a full refactor the choice of language and acceleration tools could be more carefully considered with the end goals better defined.
- Science: Many requests have been made of STEEL by others recognising it as a potentially revolutionary tool. Some are possible but hard, for example adding gas fractions requires a 4^{th} dimension to be added to the running/output data. Others are (near)impossible under the current logic of the code such as creating partitions in the central population and treating each population differently. These again could be designed into a full refactor.
- Output: The data output from STEEL and the 'post-processing' modules require significant refactoring. The simple initial models could save files by name with only 9(/18) parameter combinations that were of interest. This will not scale well and a full rethink of the data handling, model launching, and output plotting is required for significant future extensions.
- Integration: Relying on all the above STEEL has been proposed by PJG as a necessary analysis tool for ensuring the self-consistency of future extra-galactic surveys. Further details will be given in Chapter 6 and a technical outline can be found in Appendix E where a 4-year fellowship outline of work can be found.

2.6 Proof of concept

```
"""This is the original proof of concept for unevolved
SHMF creating the theoritical upper limit"""

import numpy as np
import matplotlib.pyplot as plt

import SEM # Module containing semi-empirical routines

#Subhalomass function parameters macc/M0
Unevolved = {
'gamma' : 0.22,
'alpha' : -0.91,
'beta' : 6,
'omega' : 3,
'a' : 1}

#Returns the Unevolved SHMF from Jiang, van den Bosch.
#Units are Mvir h-1
def dn_dlnX(Parameters, X):
    """
    Caculates subahlo mass funtions
    Args:
        Parameters: Dictonary containg 'gamma', 'alpha', 'beta', 'omega', 'a'
        X: m/M arrays desired subhalo/parenthalo
    Returns:
        dn_dlogX_arr: Numberdensitys per dex #N dex-1
    """
    Part1 = Parameters['gamma'] * np.power(Parameters['a'] * X,
                                           Parameters['alpha'])
    Part2 = np.exp(-Parameters['beta'] *
                   np.power(Parameters['a']*X, Parameters['omega']))
    dn_dlnX_arr = Part1*Part2
    dn_dlogX_arr = dn_dlnX_arr*2.30
    return dn_dlogX_arr # N dex-1

Binwidth_SDSS = 0.01
# Range of host halo massed to investigate linear will require weighting later
```

```
CentralHaloMass = np.arange(12+ np.log10(h),
                            15+ np.log10(h),
                            Binwidth_SDSS) #Mvir h-1

# Range of satilite masses to investigate
# starting at 10^11 as we are looking for satilites above 10^10
SatBin = 0.1
SatHaloMass = np.arange(11+ np.log10(h), 15+ np.log10(h), SatBin)

# Makes m/M as required by our Jing et al
m_M = np.array([SatHaloMass - i for i in CentralHaloMass])

# Unevolved SHMF

# Runs the model from Jing
Out = dn_dlnX(Unevolved, np.power(10, m_M))
# Weight the output to the HMF(central)
Out_Weighted = np.array([thing*HMF_fun(CentralHaloMass[i]- np.log10(h))
                         for i, thing in enumerate(Out)])

# Abundance matching
StellarX = SEM.DarkMatterToStellarMass(SatHaloMass- np.log10(h),
                                       0, Paramaters, ScatterOn = False)
# Masscuts
StellarX_10 = StellarX[StellarX > SatiliteMassCut]
# Integrates
Integrals = np.array([trapz(thing[StellarX > SatiliteMassCut], StellarX_10)
                      for thing in Out_Weighted])

# f=sum(Bin)/(sum(population)*binwidth)
AnalyticModel = Integrals/(np.sum(Integrals)*Binwidth_SDSS)

# Plotting
plt.plot(CentralHaloMass, AnalyticModel)
plt.savefig("./Figures/AnalyticMax.png")
plt.clf()

"""Proof of concept over"""
```

Chapter 3

Probing the origins of the satellite galaxy distribution in central haloes.

With more purpose than the course your life was on
But if you haven't seen the places you have come from
Then you haven't seen how far you have come
In the bigger picture, there's a star sixteen times fainter
In the bigger picture, there's a course seventeen times straighter
In the bigger picture, there's a dream eighteen times greater
And it'll steer you like no other, in the bigger picture...
Stornoway - 'The Bigger Picture'

3.1 Background

In the ΛCDM model of the universe growth of haloes is hierarchical. Haloes over time acquire a substructure of accreted haloes. Given each halo is thought to contain a galaxy at its centre the halo substructure implies that galaxies should be similarly arranged. The potential application of this is that satellite galaxies situated can, therefore, be used as tracers of dark matter and used to constrain the cosmological parameters within a given ΛCDM cosmology. However, to be a reliable method of predicting the distribution of dark matter and a good test of ΛCDM cosmology a robust model of the connection between dark matter assembly and satellite galaxies is required.

In this chapter, we explore the ability of STEEL to reproduce satellite distributions over multiple epochs. By exploiting the transparent nature of semi-empirical modelling and the flexibility of STEEL we probe the impact on satellite galaxy distributions of changing: the dynamical friction timescale, stripping of satellite stellar mass, and star formation rate. In turn, probing the strength of these effects which are crucial to understanding the nature of satellites and their mergers with central galaxies.

3.2 Halo Structure and Dynamical Friction

To constrain the evolution of the subhalo/galaxy population we need to isolate which subhalo mass bins from the unevolved subhalo accretion history have yet to finish their dynamical friction timescale and thus have survived to following epochs. The sum of all the surviving subhaloes (at each epoch) then yields the unevolved surviving subhalo mass function (USSHMF).

3.2.1 Sub-halo Merging Timescale

A key parameter used to calculate the unevolved surviving SHMF is the "observability timescale" (or survival time) of each subhalo mass bin $[M_{h,sat}(z),\, M_{h,sat}(z)+dM_{h,cent}(z)]$ associated to a parent halo mass bin $[M_{h,cent}(z),\, M_{h,cent}(z)+dM_{h,cent}(z)]$. This timescale is equivalent to the merger timescale τ_{merge} of a subhalo of mass $M_{h,sat}$ in a parent halo mass $M_{h,cent}$, which is extracted from numerical simulations and represents the (average) time a subhalo "survives" orbiting within the parent halo. To calculate τ_{merge} we use the routines in Equation 3.1 derived from N-body simulations [64],

$$\tau_{merge} = (f_{t_{dyn}}\tau_{dyn})\frac{A(M_{h,cent}/M_{h,sat})^B}{\ln(1+M_{h,cent}/M_{h,sat})}\exp\Big(C\frac{J}{J_c(E)}\Big)\Big(\frac{r_c(E)}{r_{vir}}\Big)^D, \tag{3.1}$$

where A=0.9, B=1.0, C=0.6, D=0.1 [140]. The factor τ_{dyn} is given by [105],

$$\tau_{dyn} = 1.628h^{-1}\mathrm{Gyr}\Big(\frac{\Delta_{vir}(z)}{178}\Big)^{-\frac{1}{2}}\Big(\frac{H(z)}{H_0}\Big)^{-1}. \tag{3.2}$$

Our method of considering average halo mass and accretion histories does not permit tracking single orbits and associated orbital energies. We assume instead an average orbit circularity of 0.5 [141], thus reducing the dependence on the angular momentum and radial components, $\frac{J}{J_c(E)}$ and $\frac{r_c(E)}{r_{vir}}$, to a constant. In other words, this approximation is consistent with the approach of taking the average expected orbits of subhaloes at fixed parent halo mass. The key parameter of our analysis is the factor $f_{t_{dyn}}$ included in Equation 3.1. The fudge factor $f_{t_{dyn}}$ takes into account the systematic uncertainties induced by numerical resolution effects in N-body simulations which are unable to resolve the full merging timescales of subhaloes and/or the satellite galaxies they host [112]. The parameter $f_{t_{dyn}}$ increases or decreases the dynamical times of "merging" satellites enabling an exploration of the effect of dynamical time on the final number density distributions of satellite galaxies at any given epoch.

3.2.2 Surviving Subhalo Population

At each redshift we use the unevolved subhalo accretion history and the observability timescale τ_{merge} to calculate the total 'observable' subhalo population associated with

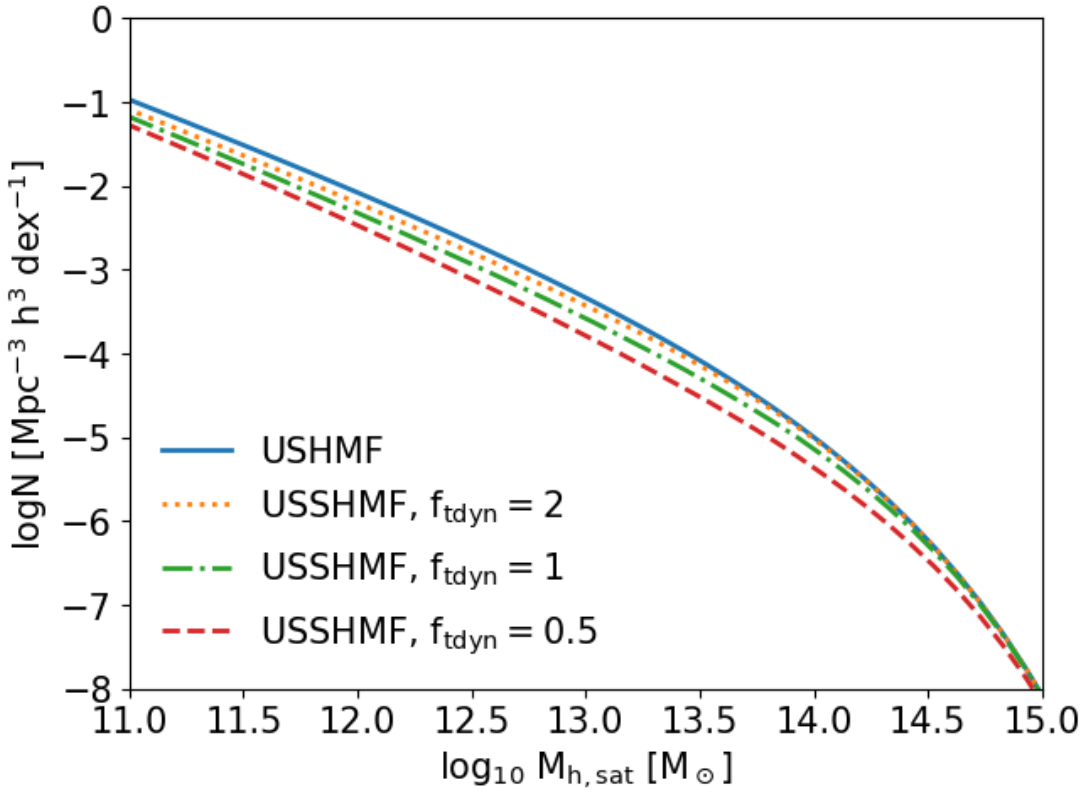


FIGURE 3.1: Example of the 'total' unevolved SHMF (solid line) and three 'total' unevolved surviving SHMF (dotted lines) corresponding to three different $f_{t_{dyn}}$ factors.

any given parent halo mass bin $[M_{h,cent}(z),\ M_{h,cent}(z) + dM_{h,cent}(z)]$, i.e. the unevolved surviving SHMF (shown by the dashed lines in Figure 2.4). To compute the implied total number densities of unmerged subhaloes with mass $[M_{h,sat}(z),\ M_{h,sat}(z) + dM_{h,sat}(z)]$ at any redshift of interest we convolve the unevolved surviving SHMF with the HMF,

$$N(M_{h,sat}, z) = \int USSHMF\left(\frac{M_{h,sat}}{M_{h,cent}}\right) HMF(M_{h,cent}, z) dM_{h,cent}. \tag{3.3}$$

Figure 3.1 shows the total observable subhalo population for different $f_{t_{dyn}}$ similar to the SHMF in Figure 2.4. Furthermore, via appropriate abundance matching algorithms, we can assign corresponding satellite galaxies to the unevolved subhalo accretion history and obtain the distribution of satellites in a given parent halo mass bin $[M_{h,cent}(z),\ M_{h,cent}(z) + dM_{h,cent}(z)]$ by assuming the satellites follow the same merging timescales as their host subhaloes.

3.3 Observed Satellite Distributions

As shown in Equations 3.1 and 3.2 the dynamical friction/observability timescale is dependent on the ratio of halo to subhalo mass. In Figure 3.2 in the left-hand panel the

merging timescales for subhaloes falling into three different central masses at redshift $z = 3$ are shown. In each case, smaller subhaloes, with lower mass ratios, have longer merging times. To probe the effect of using an average orbital circularity of 0.5 the analysis is repeated with the orbital circularity set to 0.9 and 0.1 the shading in Figure 3.2 shows the magnitude of these changes.

For subhaloes that are significantly under an order of magnitude of the mass of the central halo, the merging timescales are longer then the age of the Universe and such objects would still be present in halo structures today. Through the association of galaxies to subhaloes, via the SMHM relationship at infall, we can repeat the exercise with satellite galaxies. It is similarly found that small galaxies in massive haloes are the only galaxies accreted at redshift $z = 1.5$ that are still orbiting local galaxies.

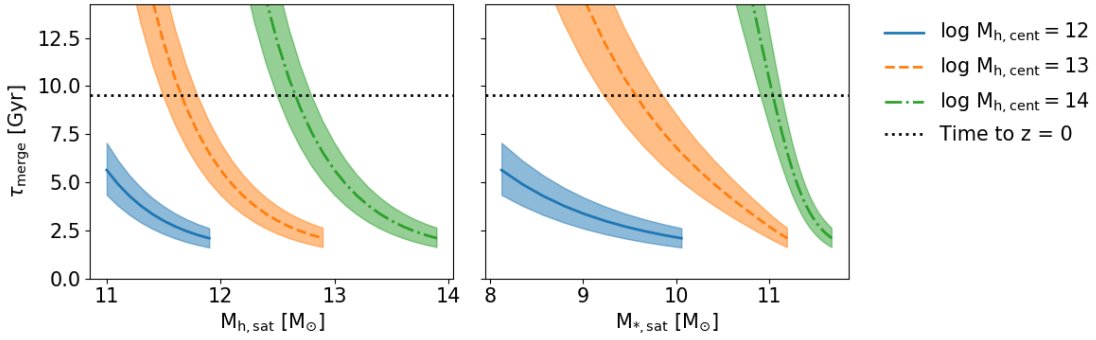


FIGURE 3.2: The range of merging timescales for a range of subhalo (left) and satellite (right) masses when accreted at $z = 1.5$ onto three different host masses: log $M_{h,cen}$ $[M_\odot] = 12$ (blue solid), 13 (orange dashed) and 14 (green dot-dashed). The shading shows the effects of changing the assumed mean orbital circularity between extreme values of 0.1 and 0.9. The dotted black line shows the time to redshift $z = 0$, i.e. the minimum amount of time a satellite would need to survive to be observable in the local universe.

In Figure 3.3 the practical effect of increasing the dynamical time is shown. Increasing the dynamical time factor f_{tdyn} to infinity, satellites will never merge with the central and thus the left-hand panel shows the total buildup of satellites over cosmic time. In the right-hand panel, with $f_{tdyn} = 1$, we see how the population of satellites still observable at $z = 0.1$ is built up over time. As expected from Figure 3.2, only recently accreted massive satellite galaxies contribute to the observable population whereas smaller galaxies will have a wider distribution of accretion times. Notably, there are no galaxies, within our mass ranges, accreted before redshift $z = 2$ that remain observable at redshift $z = 0.1$.

3.3.1 Effects of Dynamical Friction

In this Section, we show the prediction of STEEL with different merging timescales, f_{tdyn} $= 0.5, 1.0, 2.5$, to probe the effects of dynamical time on the satellite population. In Figure 3.4 we find, as expected, that longer dynamical times tend to increase satellite

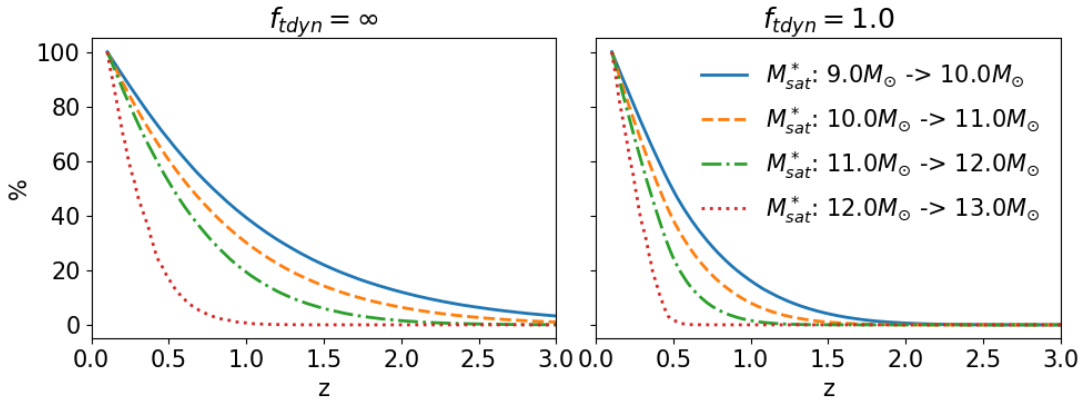


FIGURE 3.3: We show the percentage of satellites observed at a $z = 0.1$ as a function of their redshift of accretion $z > 0$. It can be seen that massive satellites observed at $z = 0.1$ are accreted more recently than smaller satellites. At $z = 0.5$ less that 50% of the total satellites observed at $z = 0.1$ have been accreted, and at $z = 0.1$ this falls to less than 20%.

number densities especially towards lower stellar masses. The number densities of higher-mass satellites are more resilient to increase with dynamical time. In fact, when $f_{tdyn} \gtrsim 1$ the number densities of massive satellites become already very close to the theoretical maximum number density when $f_{t_{dyn}} = \infty$. This trend is also seen in Figure 3.3, the difference between the buildup of high mass satellites between $f_{t_{dyn}} = \infty$ and $f_{t_{dyn}} = 1$ is much less significant than between that of low mass satellites. It follows that massive satellites are on average a recently accreted population. In other words, there are only a few high mass satellites that had not enough time to merge when $f_{tdyn} \gtrsim 1$. In contrast, lower mass satellites have not yet reached their theoretical limit, and thus they can still increase their number densities with increasing dynamical time.

When computing the distribution of satellites as a function of parent halo mass we show both full number densities, as well as fractional distributions to better highlight the "skewness" of the predicted distributions with respect to the data. The latter will be simply computed as

$$F(dM_h) = \frac{N(> 10^x)|_{dM_h}}{N(x)}, \tag{3.4}$$

where $N(> 10^x)$ is the total number (density) of satellites above a threshold stellar mass $x = \log M_*$, and the $N(> 10^x)|_{dM_h}$ is the number of these that reside within the halo mass bin $[M_h, M_h + dM_h]$.

Figure 3.5 shows how as a function of parent halo mass the distribution of satellites, above three stellar-mass cuts, is affected by f_{tdyn}. The number density distribution (top row) shows results similar to the satellite SMF where increasing dynamical time increases the number densities. However, there is also an apparent steepening effect for which lower mass host haloes end up containing relatively fewer satellites with respect to models with

longer dynamical times. The fractional plot (bottom row) accentuates this change in the number density distributions shown in the top row: shorter dynamical times shift the peak of the distribution to the right as relatively more satellites are observed in high mass host haloes.

This steepening of the satellite distribution as a function of halo mass, as well as the shift of the peak in the lower row, where infinite dynamical times move satellites preferentially to lower halo masses, are both caused by the amount of time satellites survive in their hosts. Massive satellites are far more common in massive hosts, as can be inferred from the SHMF. Therefore, irrespective of the chosen merging timescale, there will always be a high number density of surviving massive satellites in higher mass parent haloes. However, when merging timescales are increased, the lower number densities of massive satellites in moderately-sized haloes are also increased. Given that lower mass parent haloes are more abundant, the reduction of merging timescales tends to shift the peak of the fractional distribution of galaxies to lower mass parents. Otherwise put, the merging timescales in clusters ($\log_{10} M_{h,cent}\ [M_\odot] > 13.75$) are so long that even a factor five reduction in merger timescale still does not give the satellite galaxies sufficient time to merge with the central galaxy. This effect can be inferred from Figure 3.2, where we see steeper gradients for a given subhalo/satellite mass with increasing parent halo mass. Merging is more efficient for the lower mass satellites as can be seen by the steepness of the $f_{t_{dyn}} = 0.5$ model (dashed lines) in the top row of Figure 3.5.

The least-square residuals to the SMF, number density distribution and fractional distributions are given in Table 3.1. There is no model that simultaneously fully matches all the observations in all mass ranges. The trend in both the number density distribution and the fractional distribution is that slightly longer dynamical times ($f_{tdyn} = 1.2$) are favoured by the less massive satellites. Longer dynamical timescales better match the halo mass distributions (number density distribution/fractional distributions) for lower mass satellites, and vice versa for higher mass satellites with $\log M_*/M_\odot > 11$. Nevertheless, the simple combination of abundance matching and dynamical merging timescales as suggested by pure N-body simulations ($f_{t_{dyn}} = 1.0$) tends to provide overall good agreement to both the satellite SMF and the satellite distributions, without the need to invoke additional physics in the (late) evolution of satellite galaxies after infall.

3.3.2 Evolutionary models

The prior section makes an implicit assumption that satellites do not evolve after infall. In this Section, we apply several evolutionary effects to the satellites assuming they have the same properties as centrals at infall but then evolve due to in-situ processes and in response to their environment.

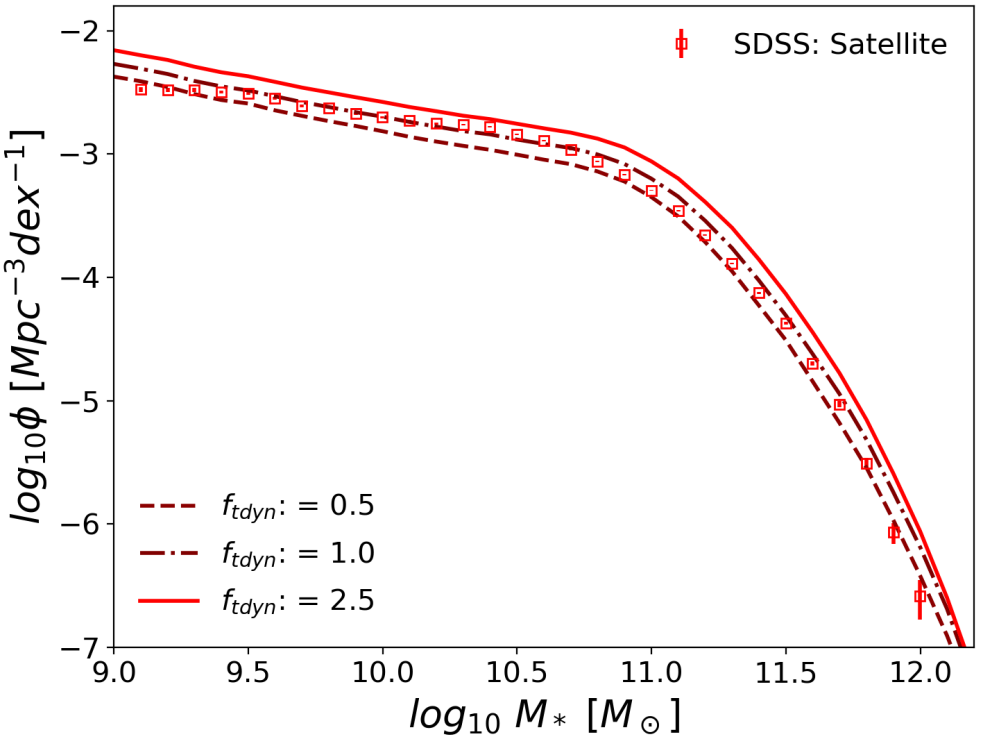


FIGURE 3.4: Satellite stellar mass functions generated by the model compared to SDSS data (open squares). The solid, dot dashed, and dashed lines show $f_{tdyn} = 0.5, 1.0$, and 2.5 respectively.

TABLE 3.1: We show the sum of the squared residuals between the SDSS and our model. The satellite SMF is calculated between 9.1 and 12.0 M_*. The SDF fit is calculated between 12 and 14.9 M_h for the >10 and>10.5 plots, and between 12.5 and 14.9 M_h for >11. The Fractional plot fit is calculated between 11.6 and 14.9 M_h.

$f_{t_{dyn}}$	SSMF (Fig 3.4) (Fig 3.4)	SDF (Top Row Fig 3.5)			Fractional Distribution (Bottom Row Fig 3.5)		
		>10	>10.5	>11	>10	>10.5	>11
0.5	0.022	0.19	0.55	0.073	0.0042	0.0047	0.0078
0.8	0.025	0.13	0.51	0.089	0.0020	0.0017	0.0054
1.0	0.034	0.12	0.56	0.10	0.0015	0.0011	0.0050
1.2	0.043	0.12	0.53	0.11	0.0015	0.00094	0.0046
1.5	0.054	0.12	0.52	0.13	0.0017	0.0010	0.0045

3.3.2.1 Stripping

The stellar stripping is implemented following the empirically-based formalism suggested by Cattaneo et al. [94]. Satellite galaxies strip stellar mass proportionally to the ratio of the host subhalo and the parent halo

$$M_{*,sat} = M_{*,sat}(1 - \eta_{strip})^{\tau_{strip}}, \tag{3.5}$$

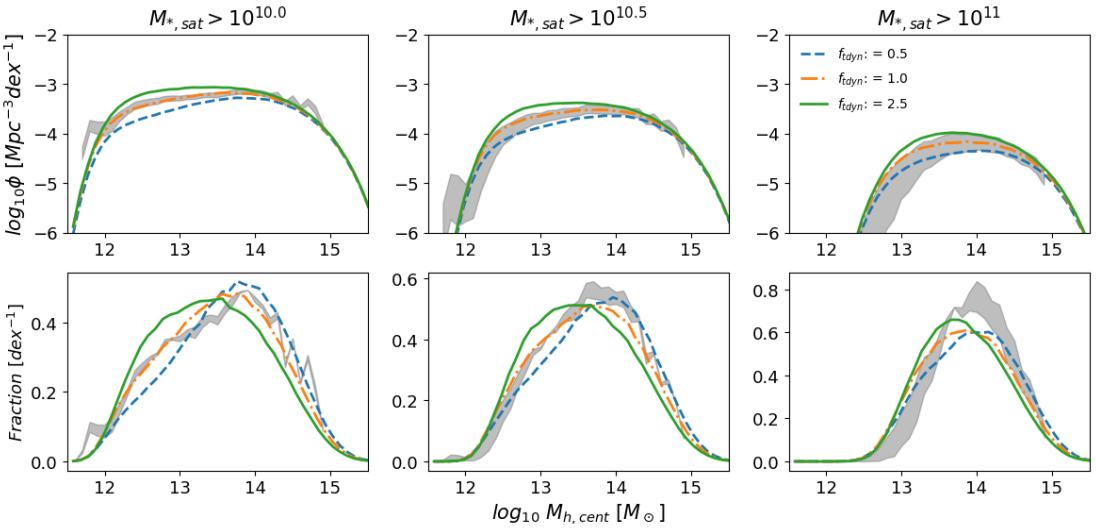


FIGURE 3.5: Satellite distributions in parent haloes generated from STEEL are compared to those observed in SDSS (grey band). Columns from left to right show increasing satellite stellar mass cuts as labelled. The top row shows the number density of satellites expected to be found in each parent halo mass. The bottom row shows the fractional distribution described by Equation 3.4. The dashed, dot dashed, and solid lines show $f_{tdyn} = 0.5, 1.0$, and 2.5 respectively.

where τ_{strip} is the dynamical friction timescale in orbital time. For our model, where the orbital circularity is assumed to take the average value, this becomes the dynamical friction time [63] times a constant given by [142] using the average orbital circularity. Following Cattaneo et al. [94] the latter can be written,

$$\tau_{strip} = \frac{1.428}{2\pi} \frac{M_{h,cent}/M_{h,sat}}{\ln(1 + M_{h,cent}/M_{h,sat})}. \tag{3.6}$$

We set as a reference $\eta_{strip} = 0.4$ as suggested by Cattaneo et al. [94] who showed that larger values would be inconsistent with the observational constraints in the local universe. If at the time of observation, say $z = 0$, a galaxy's full dynamical time is not yet passed, a time-dependent reduction factor is applied to the amount of total stellar stripping given by Equation 3.6.

3.3.2.2 Star Formation Rates

We use the star formation rate (SFR) parameterization from Lee et al. [143] with parameters[1] from Tomczak et al. [11] where s_0 and M_0 have units $log(M_\odot)$ and $M_\odot$ respectively,

[1]These parameters are derived by fitting data from ZFORGE in combination with far-IR imaging from *Spitzer* and *Herschel* in the range $0.5 < z < 4$. In this work we extrapolate their fits down to $z = 0$, as this is consistent with the SFR measured by [144] at lower redshifts.

$$\begin{aligned} \log[\psi(z, M_*)] &= s_0(z) - \log\left[1 + \left(\frac{M_*}{M_0(z)}\right)^{-\alpha(z)}\right] \\ s_0(z) &= 0.195 + 1.157z - 0.143(z^2) \\ \log[M_0(z)] &= 9.244 + 0.753z - 0.090(z^2) \\ \alpha(z) &= 1.118. \end{aligned} \tag{3.7}$$

As discussed, continuity-equation approaches are more realistic and physically-based techniques to model galaxy growth, as, by design, they avoid the observed discrepancies between the time-integrated SFR and the independently measured stellar mass function at any given epoch. The novelty in our continuity model with respect to previous work is that we do not tune the resulting star formation rate on the total stellar mass function but rather only on the stellar mass function of *central* galaxies [2]. We use as input for our continuity equation the central SMF generated by the SMHM relation and the central HMF from [145]. When considering the mass growth we must consider the mass loss, else the SFR is massively under-predicted. We use an instantaneous loss fraction of 40%. We also neglect the mass gained from mergers[3]. However, as mergers would decrease the SFR calculated, this method should be considered an upper limit to the real SFR. The fit to the resulting SFRs is given by

$$\begin{aligned} \log(\psi(z, M_*)) &= s_0(z) - \log\left[1 + \left(\frac{M_*}{M_0(z)}\right)^{-\alpha(z)}\right] \\ s_0(z) &= 0.6 + 1.22z - 0.2(z^2) \\ \log(M_0(z)) &= 10.3 + 0.753z - 0.15(z^2) \\ \alpha(z) &= 1.3 - 0.1z. \end{aligned} \tag{3.8}$$

In all cases, the SFR is included in our models with a log-normal scatter of 0.3 dex [83]. We also include the ability to quench star formation in satellite galaxies after infall. It has been suggested that satellites undergo a "delayed-then-rapid" quenching [9]. The latter model envisions that satellites continue to form stars at the same rate as central galaxies of comparable stellar mass for a time τ_q after infall, and then quench rapidly over a timescale τ_f. This quenching is proposed for satellites with stellar mass above $M_* \gtrsim 10^9 M_\odot$, with a minimum $\tau_q \sim 1\ Gyr$. For galaxies below $M_* \lesssim 10^9 M_\odot$, we however adopt the more recent results by Fillingham et al. [10], who put forward a parent halo

[2] Which is in turn iteratively constrained by matching the local stellar mass function of SDSS centrals.

[3] This mass loss and neglecting mergers is used for Equation 3.8, the integrated mass loss and mergers as described in Chapter 2 are used in the starformation rate used in Chapter 4.

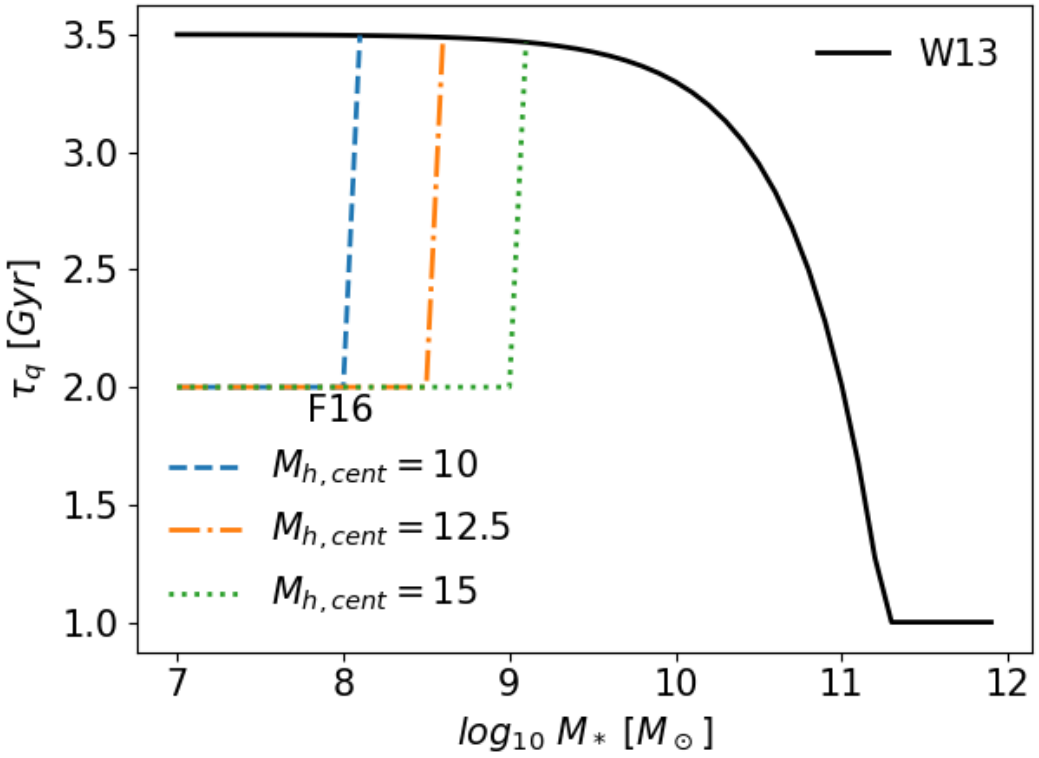


FIGURE 3.6: The solid line shows the Wetzel et al. [9, W13] model for quenching. The dashed lines show the host halo dependent reduction in quenching time from Fillingham et al. [10, F16] for three example host masses $\log 10\ M_{h,cent} = 10, 12.5, 15$ as labelled. Larger hosts are able to reduce the quenching time of larger satellites

.

mass dependent cutoff,

$$log(M_{cutoff}) = 9log(M_\odot) - (15log(M_\odot) - log(M_{h,host}))/5log(M_\odot), \tag{3.9}$$

below which satellite galaxies all share the same quenching time $\tau_q = 2$ Gyr. The rapid quenching timescale τ_f can be expressed as [9],

$$\tau_f = -0.5log(M_{*,sat}) + 5.7Gyr. \tag{3.10}$$

We set a minimum τ_f of 0.2 Gyr for all galaxies. This rapid quenching begins at times $t > \tau_q$ after infall, and it is approximated by an exponential decay which is longer for larger satellites, as can be inferred from Equation 3.11. The look-back time at which a galaxy begins fast quenching is then $t_q = t_{infall} - \tau_q$. After this time the satellite no longer follows the SFR of a typical central galaxy. Figure 3.6 illustrates the quenching model where the dashed/coloured lines mark the halo mass dependence in cutoff mass. The SFR during the satellite infall is then given by,

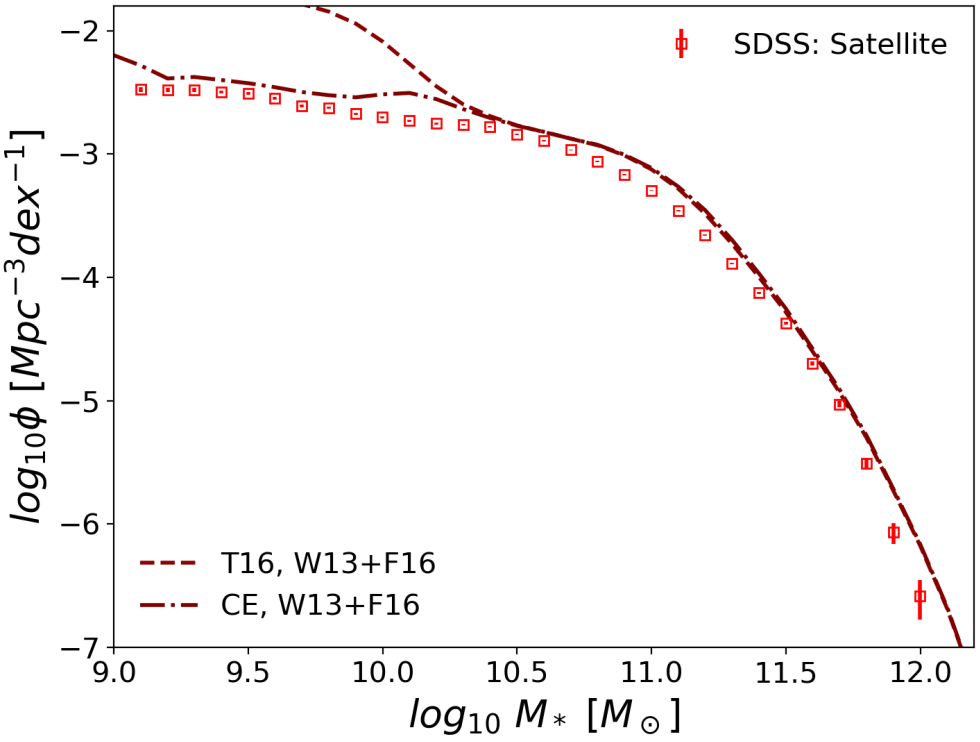


FIGURE 3.7: Satellite stellar mass functions generated from the model using both the Tomczak et al. [11] (dashed) and continuity (dot dashed) star formation rates compared to the SDSS satellite stellar mass function (open squares).

$$SFR(t, M_*) = SFR(t, M_*) \begin{cases} \psi(z(t), M_*), & t > t_q \\ \psi(z(t_q), M_*) e^{\left[-\frac{t_q - t}{\tau_f}\right]}. & t < t_q \end{cases} \tag{3.11}$$

If at any point a satellite galaxy has a SSFR below $10^{-12}\ M_\odot\ yr^{-1}$, it is assumed to be fully quenched and assigned a SSFR of $10^{-12}\ M_\odot\ yr^{-1}$, plus a log-normal scatter of 0.3 dex.

The first, purely observationally-based, star formation rate model strictly follows the SFR parametrization by Tomczak et al. [11, T16 hereafter], given in Equation 3.7. The second star formation rate model (which we label as "CE") is instead based on the continuity equation approach.

We compare the satellite SMFs produced by the two star formation+quenching models addressed above to our SDSS stellar mass function of satellites in Figure 3.7. It is apparent that using the observed SFR by T16 (dashed line), even inclusive of the best recipes for quenching, still substantially overproduces the number density of galaxies below $M_* \lesssim 3 \times 10^{10}\ M_\odot$. This is a well-known problem affecting the full (dominated by central) galaxy population [e.g., 83]: the integrated (observed) SFR is not consistent with the moderate growth over time of the SMF causing an overproduction of galaxies

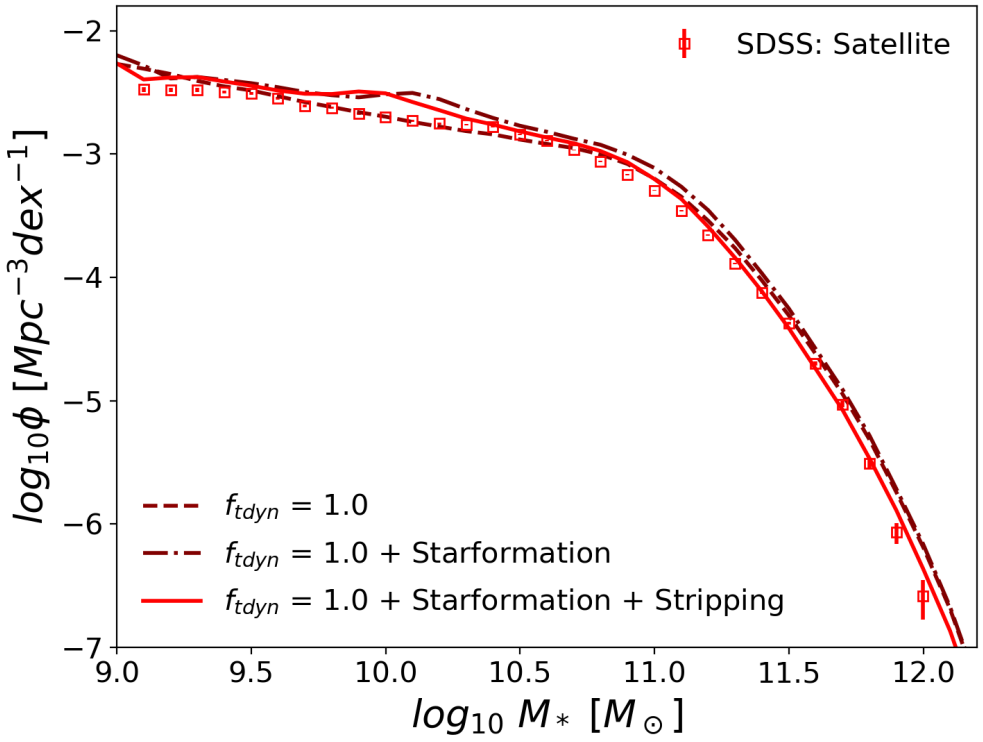


FIGURE 3.8: Satellite stellar mass functions generated from the model compared to SDSS satellites (open squares). The models shown all have $f_{tdyn} = 1.0$ and are the reference 'frozen model' (dashed line), starformation (CE model) only (dot dashed line) and starformation and stripping (solid line).

becoming gradually more severe at lower stellar masses. Our results show a similar problem affecting the satellite population, on the assumption that the latter at infall share the same SFR distribution as a typical central galaxy of the same stellar mass.

We now show the relative impact of the star formation rate and stellar stripping on the satellite stellar mass function. Figure 3.8 and Figure 3.9 show stellar mass function and host halo mass distributions for the $f_{tdyn} = 1.0$ reference model with no stripping nor star-formation, the CE star formation model, and the CE star formation model with stripping (long-dashed, dot-dashed, and solid lines, respectively). We see from both Figures that the reference and star-formation model are almost indistinguishable. The stripping, at least at the level implemented in this work, also has a rather minor effect, at the most reducing the number densities of the most massive satellites ($> 10^{11} M_\odot$) by $\lesssim 0.2$ dex.

Table 3.2 shows the sum of the square residuals to the SMF, SDF and fractional distributions for the same models discussed above, our reference frozen one, and the one with evolution of satellites after infall (stellar stripping and continuity equation-based star

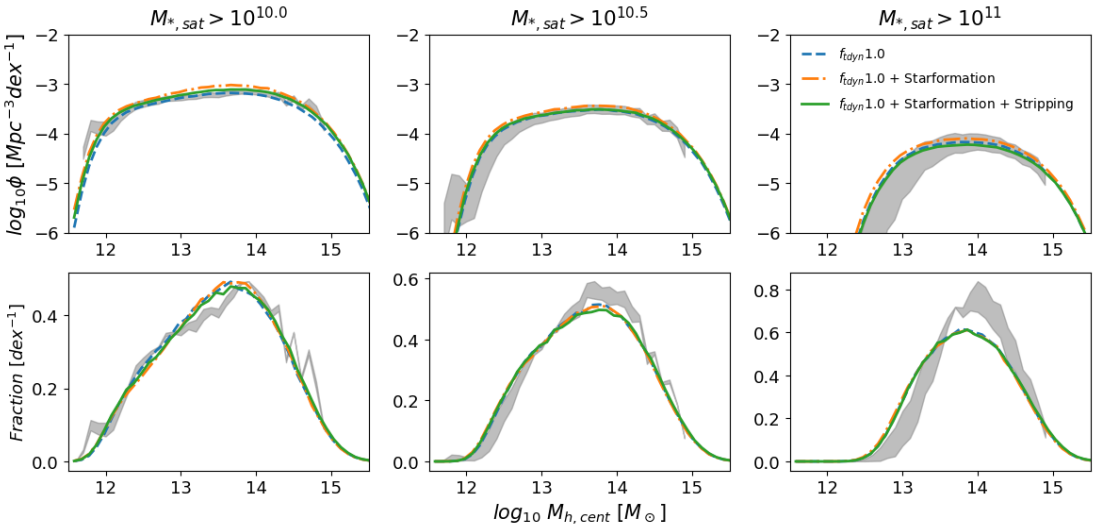


FIGURE 3.9: Satellite distributions in parent haloes generated from the model are compared to those observed in SDSS (grey bands). Columns from left to right show increasing satellite stellar mass cuts as labelled. The top row shows the number density of satellites expected to be found in each parent halo mass. The bottom row shows the fractional distribution described by Equation 3.4. The models shown all have $f_{tdyn} = 1.0$ and are the reference 'frozen model' (dashed line), star formation only (dot dashed line) and star formation and stripping (solid line). The width of the grey band corresponds to a 10% uncertainty in satellite stellar masses.

TABLE 3.2: We show the sum of the squared residuals between the SDSS and our model as in 3.1 with the same mass ranges for the fitting. All models have $f_{t_{dyn}} = 1.0$ from top to bottom we then have the reference frozen model, the model with starformation, and the model with stripping and starformation.

$f_{t_{dyn}}$	SSMF (Fig 3.8)	SDF (Top Row Fig 3.9)			Fractional Distribution (Bottom Row Fig 3.9)		
		>10	>10.5	>11	>10	>10.5	>11
1.0	0.034	0.12	0.53	0.10	0.0015	0.0011	0.0049
1.0 With Star Formation	0.049	0.077	0.43	0.14	0.0018	0.0012	0.0059
1.0 With Stripping and Star Formation	0.021	0.087	0.47	0.088	0.0016	0.0015	0.0056

formation). Table 3.2 shows that the satellite late evolution has little effect on the fractional distribution. In the number density distribution we see an improved fit for galaxies in the $M_* > 10^{11}\, M_\odot$ range, mainly induced by the stripping which slightly reduces the number density of massive galaxies. Table 3.2 shows the sum of square residuals for the dynamical time with $f_{t_{dyn}} = 1$, for the frozen and evolved models.

3.4 Multi-Epoch Distributions of Satellite Galaxies

We here extend the group and cluster satellite richness analysis to high redshift. In Section 3.3 it is found that dynamical friction and, to a second-order, abundance matching,

are the dominant factors in the distribution of satellite galaxies in groups and clusters above $M_{*,sat} > 10^{10} M_{\odot}$. Here we display the results for the full STEEL model which includes star formation, dynamical quenching and stripping to evolve satellites after infall. The latter effects, despite being of lower order than dynamical friction or abundance matching, are included to be able to compare to data other than cluster richness, such as the satellite specific star formation rate distribution.

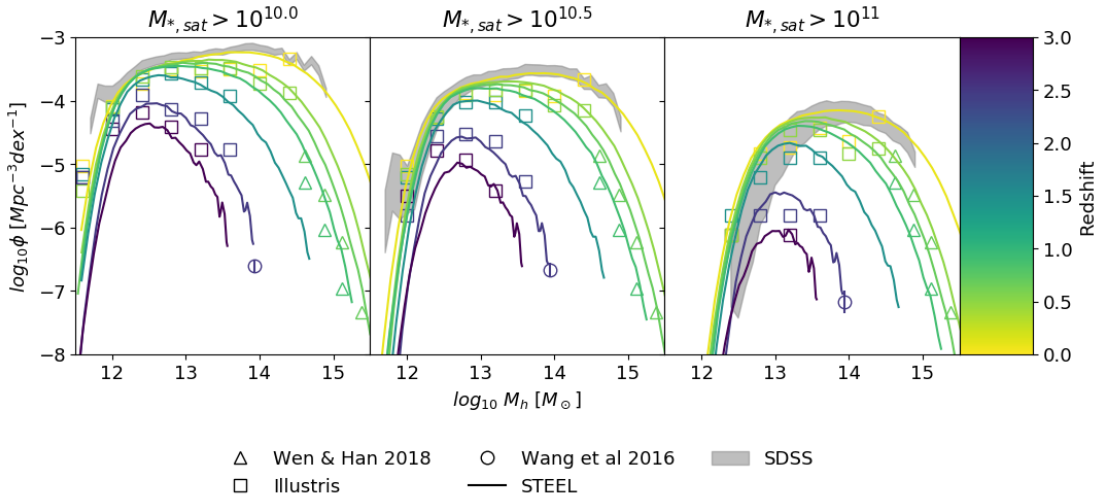


FIGURE 3.10: The number-density distribution of satellites per parent halo mass predicted from STEEL, using the PyMorph SMHM relation, at multiple redshift epochs (solid lines). The grey band is the data from SDSS at redshift $z = 0.1$. Also included are the high redshift cluster data from Wang et al. [12] (circles) and Wen and Han [13] (triangles). For completeness, the Figure also includes the outputs from the Illustris simulation using the TNG100 data (crosses). Each data point and line are given a colour associated to their redshift (the bar on the right provides the color coding key).

Figure 3.10 shows the satellite number density per halo mass bin. For each central halo mass the cosmic number density, similar to the number density presented in the cumulative stellar mass functions, is calculated for satellites above a mass threshold for each halo mass bin. The predicted halo richness from STEEL, using the PyMorph SMHM relation, is shown in this plot as solid lines. The predictions from the Illustris TNG100 simulation [146, 147] are shown with crosses. Low redshift SDSS data are shown as a grey band, cluster data detailed in Section 2.4.1.2 are open symbols. The markers and lines in the figure are colour coded based on redshift, as indicated by the colour bar on the right.

3.5 Discussion

There is a vast literature on the modelling of satellite galaxies. Here we recall examples of semi-empirical and semi-analytic models to highlight some of the key similarities and key differences. Neistein et al. [148], with an approach similar to ours, separated the central

and satellite populations in an attempt to better define the galaxy halo connection. By allowing in an N-body simulation the stellar mass of satellite galaxies to depend on both the host subhalo mass and the parent halo mass, Neistein et al. [148] find that the local satellite stellar mass-halo mass is substantially less well defined than the one for central galaxies. In our model satellites instead, strictly follow the stellar-mass-halo-mass relation of centrals at infall. In this way we find the resulting satellite distributions to be well reproduced. Our semi-empirical statistical model was able to reproduce multiple observables such as the stellar and parent halo mass distributions, with essentially only one parameter, $f_{t_{dyn}}$. By working with minimal assumptions and related free parameters, our approach is thus less prone to possible degeneracies affecting more traditional, multi-parameter techniques.

Another key difference with respect to previous models concerns "orphan galaxies". In N-body or merger tree-based simulations, when a subhalo drops below the resolution limit, an orphan galaxy is created [e.g., 31, 58]. It is then necessary to make an assumption on how much longer that subhalo (and hosted satellite galaxy) will survive. In our model, we avoid this complication by self-consistently assigning to all satellites a (full) observability timescale at infall.

We found in Figure 3.10 that STEEL is able to predict the existence of extreme objects. It is able to use rich cluster environments that are observable up to high redshift and contain some of the most massive galaxies as comparative data. This gives STEEL an edge as exploring the richness of the environments around massive galaxies provides an excellent constraint to hierarchical assembly predicted by ΛCDM cosmology at the most extreme masses [96]. For example, we successfully compare with the cluster reported in Wang et al. [12], which other models [e.g. 149] have been unable to reproduce within a ΛCDM framework. However we concur with Wang et al. [12] that these objects are rare (i.e. low number density) and their absence in traditional simulations could be simply attributed to poor statistics (i.e. small volumes) and not necessarily to the implied physical model. With large-scale surveys such as EUCLID coming online, a well-tuned statistical model could more easily place robust constraints on high-redshift cluster formation.

In the following chapters we will use these tight constraints on the satellite distributions, and by extension satellite accretion rates, in conjunction with the statistical halo growth histories to test the total mass assembly statistics of galaxy populations. This is a departure from classical models which collect the assembly statistics for individual galaxies then take a population average. Additionally, we are able to use the flexible nature of a semi-empirical model to begin to look at how the SMHM relation can be used to understand systematic differences introduced by assumptions behind data analysis algorithms (such as M/L, IMF, etc...). The methodology presented in Chapter 2 and the results in this chapter open the gates to novel analysis providing a unique view of the assembly of galaxies.

3.5.1 Future work

The quality of data presented in Figure 3.10, SDSS and targeted cluster observations is a limitation to our ability to constrain the environments of galaxies over many epochs. The data limits the technique due to the lack of halo mass estimations and satellite/central identification in the high redshift data sets. To obtain the required data products accurate clustering analysis similar to that of Yang et al. [118] must be applied to comprehensive high redshift surveys such as EUCLID.

When comprehensive high redshift data sets with satellite/central identification are available, STEEL will be well-positioned to place constraints on several aspects of hierarchical galaxy formation. Firstly, with good central SMF, the quality of abundance matching will improve significantly with the uncertainties regarding the contributions from satellite galaxies removed (Section 2.4.2). The analysis of the impact of dynamical friction timescales on the exact shape of the satellite distributions such as that shown in Figure 3.5 can be extended to high redshift. Constraints on the dynamical friction timescales will improve estimations of the central galaxy accretion and merger rates that are further discussed in Chapters 4 & 5.

Chapter 4

How do Galaxies Acquire Mass? Assembly vs. Star Formation

"All models are wrong, but some are useful." George Box

4.1 Background

Galaxies acquire stellar mass in two ways, star formation and satellite accretion (mergers). Star formation is the process of gas collapse to form stars. Galaxies at high redshift are thought to produce most of their mass through star formation. Star formation then reaches a peak at redshift $z = 2$. The cessation (quenching) of star formation remains an open question. The leading theories involve several internal and external processes, from stellar and active galactic nuclei feedback to host halo and/or morphological quenching [150–153].

In contrast to star formation, the assembly of mass though mergers is thought to increase at lower redshifts. In particular, in very massive galaxies, growth via satellite accretion has been claimed to become progressively more relevant [30, 96, 154–158]. The characteristic mass at which galaxies transition from being in-situ to ex-situ growth dominated has previously been found at $M_* \sim 10^{11} M_\odot$ [94, 155, 159]. Central galaxies that reside at the centre of massive haloes thus provide a window into the different pathways that have contributed to the mass growth history of galaxies in the local universe. Exploring the way these galaxies build their mass can give insights into the stellar-mass-halo-mass (SMHM hereafter) relation, the efficiency of the satellite transport from the edge of the cluster to the centre, the balance of the major processes taking place on these satellites, the galaxy merger rate, and the star formation rate.

4.1.1 Previous techniques

Models of galaxy formation traditionally use the hierarchical growth of dark matter structure as the backbone for galaxy assembly. Hydrodynamical simulations co-evolve the dark matter and baryonic matter allowing for a simultaneous look at the assembly of both components [27, 160]. The latter technique, however, requires large computational resources. Less computationally intensive models such as traditional Semi-analytic and Semi-empirical models, use dark matter merger trees from post-processing of dark matter simulations [31, 161]. Dark matter merger trees visualise dark matter assembly as a central trunk and halo mergers happen where branches join. Semi-analytic models initialise gas at high redshift and use several physical assumptions and free parameters to tune to observations [30, 31]. Semi-empirical models use a more direct approach initializing galaxy stellar mass in dark matter haloes most commonly through abundance matching, the association of galaxies to dark matter host haloes via relative abundances [8, 17, 32–34]. Both Semi-Empirical and Semi-Analytic models follow the merging histories of the underlying dark matter merger trees to track the in-situ and ex-situ buildup of galaxy mass. The work of Moster et al. [17], for example, uses a semi-empirical model to associate the growth of the dark matter halo to the star formation rate of the host galaxy alongside the build-up of stellar mass from satellites accretion, further strengthening the connection between the dark matter host environment and the build-up of galactic stellar mass.

It is of relevance to the calculation of in-situ vs ex-situ mass buildup that the observed star formation rate is significantly higher than continuity estimates of the star formation rate [e.g. 83, 84]. It is consequently found that if observed star formation rates are used in models, they cannot be reconciled with the stellar mass functions. This is a particular problem for semi-empirical models where one would ideally use the observed star formation rate as an input. To overcome the inconsistencies between observed star formation rates and model predictions it is possible to include continuity star formation rates.

4.2 Constraining the In-Situ vs. Ex-Situ growth in STEEL

To properly constrain the formation of a galaxy one must reproduce the galaxy environment, i.e. the distribution of satellites around the central galaxy at all previous redshifts. Discrepancies with observations of the high redshift environment will result in a deviation from the true satellite stellar mass accretion rate, to account for the satellite excess or deficit. For example, a high redshift environment with too many satellites will then have an overabundance of mergers resulting in a satellite accretion rate that is too high. To

account for the deficit/excess in stellar mass accretion the modelled in-situ growth will attempt to compensate. The compensation will be found in other modelling routines which will attempt to maintain a good match with the observed evolution of the stellar mass density. Such compensation could, for example, be of the form of suppressed/enhanced star formation rate or any number of other physical modelling parameters. Reproducing the number density and distribution of galaxies has however proven to be a challenge for many semi-analytic models [e.g. 71]. Furthermore, semi-analytic models usually characterised by a large number of modelling parameters can suffer from degeneracies severely limits our understanding of the essential physical processes governing galaxy formation [e.g. 80, 162].

4.3 Incompatible ΛCDM and Stellar Mass Functions

The cartoon in Figure 4.1 shows a simple visualisation of the process we use to determine the effect of different stellar mass functions/SMHM relations on the accretion histories and thus ex-situ growth of galaxy populations. Starting from the left panel we show two stellar mass functions, the primary difference is the blue (dotted) stellar mass function has a substantially enhanced high mass end. In the middle panel, we show how this high mass slope propagates to the SMHM relation, produces an enhanced high mass slope. The galaxy growth histories, shown as solid lines, are generated using the SMHM relation and a halo mass growth history from the statistical dark matter accretion history. Galaxy growth history calculated using the steeper SMHM relation induces more galaxy growth. Whereas, the flatter SMHM relation induces less galaxy growth. In summary, a flatter SMHM relation induces less central galaxy growth with increasing halo mass.

In Figure 4.2 we report the satellite accretion generated using the 'cmodel' SMHM relation/SMF given in Section 2.4.2. The top row shows the total mass of the galaxy and the total satellite accretion. The middle row shows the fractional contribution from satellite accretion from $z = 3$. The bottom row shows the instantaneous mass growth from satellite accretion. In Figure 4.2 we obtain a lower limit for the accretion rate by including stripping but not star-formation in the satellites, thus minimising their mass through environmental processes. We find for the high mass galaxies, which are above the knee of the SMHM relation, even the lower limit for the accretion has an instantaneous rate greater than the growth rate of the galaxy as seen in the bottom row. This makes the cmodel SMHM relation used within our dark matter accretion model *non-physical*.

Similarly, the relative contributions to the average stellar mass growth of central galaxies from satellites and star formation history are calculated from STEEL, using the PyMorph SMHM relation/SMF given in Section 2.4.2. This is shown in Figure 4.3 for three galaxy mass bins (10^{11},$10^{11.5}$,10^{12} $M_{\odot}$) selected at $z = 0.1$. The average growth history (total, solid lines) is derived by following the host halo-mass track, and the stellar-mass track

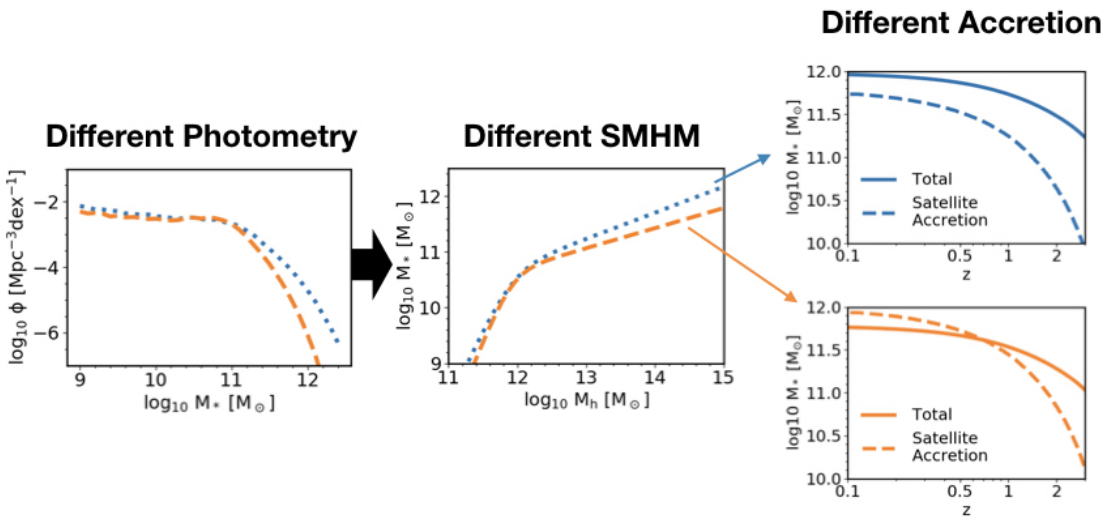


FIGURE 4.1: A cartoon showing the steps we follow to connect the differences found in the stellar mass function (left) and the changes in the SMHM relation (SMHM, middle), propagate into changes in the accretion histories (right). In the right-hand panel, dashed lines are mass from satellite accretion and solid lines are total galaxy mass growth. Flatter SMHM relations imply a weaker growth of stellar mass in the central which can be easily overcome by the substantial cumulative growth of merging satellites, rendering the model internally inconsistent.

is implied by imposing abundance matching at all redshifts. The stellar mass history assigned by abundance matching is naturally independent of any galaxy merger modelling assumptions from STEEL. The total accretion from satellites (accretion, dashed lines) is computed from the expected satellite accretion along halo mass tracks. For each galaxy, a star formation history (SFH, dotted lines) may then be calculated. The star formation rate is tuned such that it provides the correct star formation history to account for the difference between the mass growth expected from abundance matching and the cumulative satellite stellar mass accretion, as explained in detail in Section 4.4.

In Figure 4.3 we identify the epoch after which a galaxy transitions into a merger-dominated state under two definitions. Firstly, we define the "cumulative transition" epoch the time at which the galaxy has accreted more mass than it has created from star-formation processes (Points A & B). Secondly, we define the "instantaneous transition" as the epoch when the growth rate from mergers overtakes the growth rate from star-formation (Points C & D). More massive galaxies transition earlier to merger dominated growth under both definitions. However, all galaxies transition earlier under the second (instantaneous) definition. The masses shown in Figure 4.3 display three relevant galaxy accretion tracks. The $M_*^{z=0} = 10^{12} M_\odot$ galaxy growth curve at low redshift is always dominated by satellite accretion. In the top and middle rows, we see that more mass has been accreted than produced by star formation, and in the bottom row we see the accretion rate overtook the star formation rate at redshift $z = 2$. The $M_*^{z=0} = 10^{11.5} M_\odot$ galaxy growth curve has more mass created from star formation than satellite accretion.

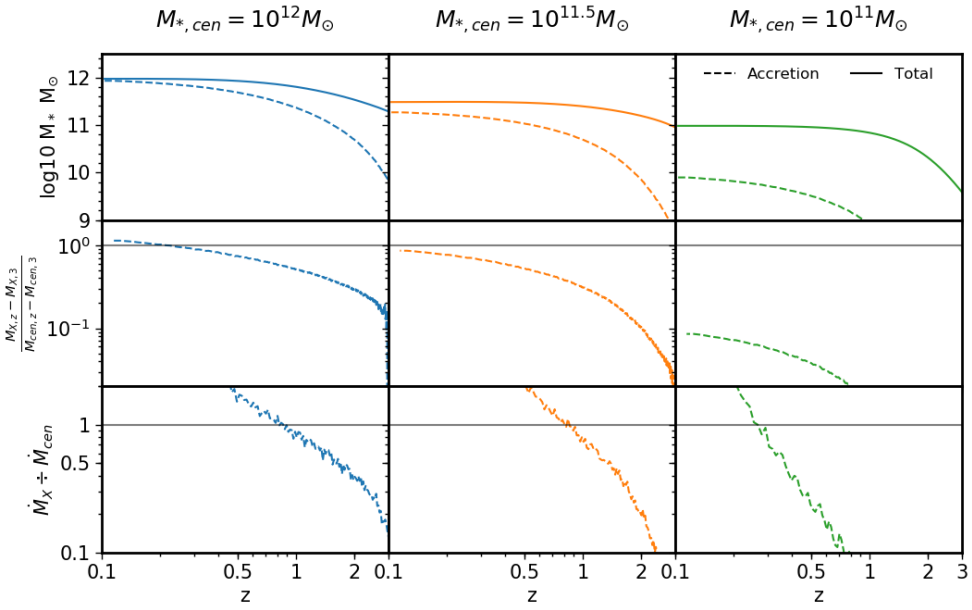


FIGURE 4.2: Three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ in blue orange and green respectively. It is clear that this model is internally nonphysical as the accretion via satellites (dashed lines) rapidly overshoots the total growth in stellar mass (solid lines) implied by the underlying growth host halo growth, as evident in the middle and bottom rows.

However, the galaxy population has a higher rate of accretion rate than star formation rate since redshift $z = 1$. The final population shown at $M_*^{z=0} = 10^{11} M_\odot$ is star formation dominated under both cumulative and instantaneous definitions. At redshift $z = 0$ we find the transition masses for the total mass ratio and the instantaneous ratio to be at $M_* = 10^{11.7} M_\odot$ and $M_* = 10^{11.1} M_\odot$ respectively.

4.4 Deriving the Star Formation Rate

The cartoon in Figure 4.4 shows the processes we adopt to derive the star formation rate by following galaxy populations along their halo mass histories.[1] The plot labelled 1 (green) is the input stellar mass function. The box in red is the statistical dark matter accretion history described in Section 2.3.1, including the halo mass function (2a), the central growth histories (2b), and the halo substructure (2c) shown here as a discrete merger tree for visualization purposes. Using the abundance matching routines described in Section 2.4.2, the stellar mass function (1) and the halo mass function (2a) are used to create the SMHM relationship (3, black).

In Chapter 3 we showed how the dark matter accretion histories (2) and abundance matching (3) can be used to generate distributions of satellites for any central halo at

[1] The coloured text is matched to the colours in Figure 4.4 intended to guide the reader through the multi-step process.

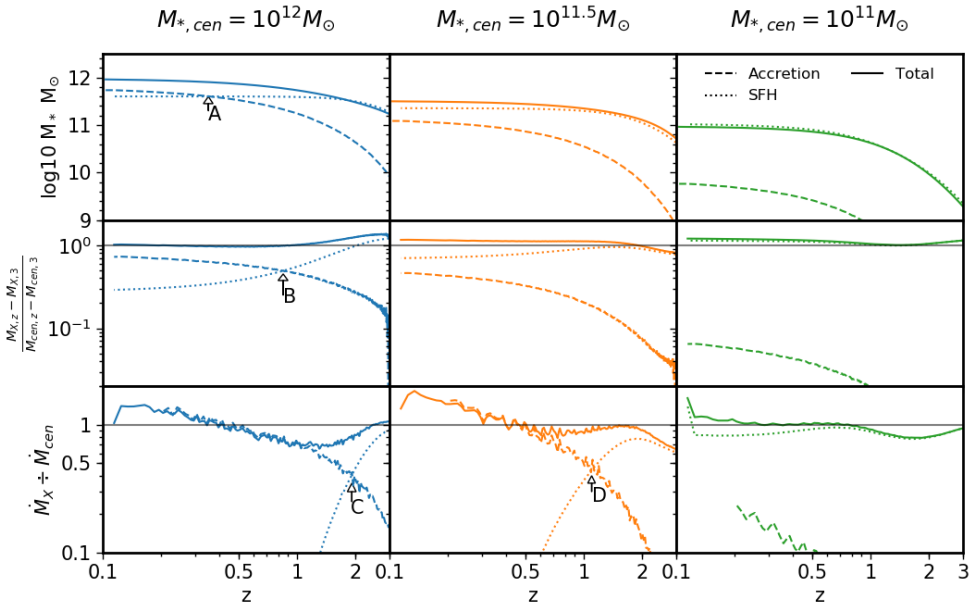


FIGURE 4.3: Three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ in blue orange and green respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panels show the total mass of the central (solid lines) and the total mass gained from accretion or star formation. The middle panels show the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panels show the ratio of the mass accretion rate from satellite galaxies, the star formation rate, and the mass growth rate of the central galaxy predicted by abundance matching. The black horizontal lines in the second and third rows are at unity. The solid lines showing the sum of the other two factors should be close to or on the unity lines. The labels A & B point to where the cumulative mass from accretion overtakes the cumulative mass from star formation. The labels C & D point to where the instantaneous accretion overtakes the star formation rate.

multiple redshifts (4). For each central halo mass track (2b) we calculate the average number density of satellites that reach the centre of the halo and merge with the central galaxy thus generating the average satellite accretion history (dashed line, 5). Using the central halo growth histories (2b) and the SMHM relation (3), we can generate the average central galaxy growth history (solid line, 5). These two quantities can be compared to check for self-consistency, as described above and shown in Figure 4.1. Where a self-consistent central growth and accretion history is found, any deficit between the accreted mass and the growth history is attributed to star formation rate (delta, 5). The derived star formation rate for central galaxies (solid line, 6) is compared to observational data (points, 6).

The star formation rate prediction generated in this way is independent of star formation observations. It relies on the (statistical) dark matter accretion history and the SMHM relationship, it is, therefore, most similar to a continuity star formation rate. The advantages of this approach compared to the continuity approach are twofold. Firstly, by

following the halo growth histories we capture the 'natural' growth of the galaxy which is an improvement to tracking galaxies by abundance between SMF epochs. Secondly, each galaxy growth history has an associated satellite accretion that is a core part of the model, whereas continuity approaches traditionally need to post process the result to include mass from accretion which is less accurate.

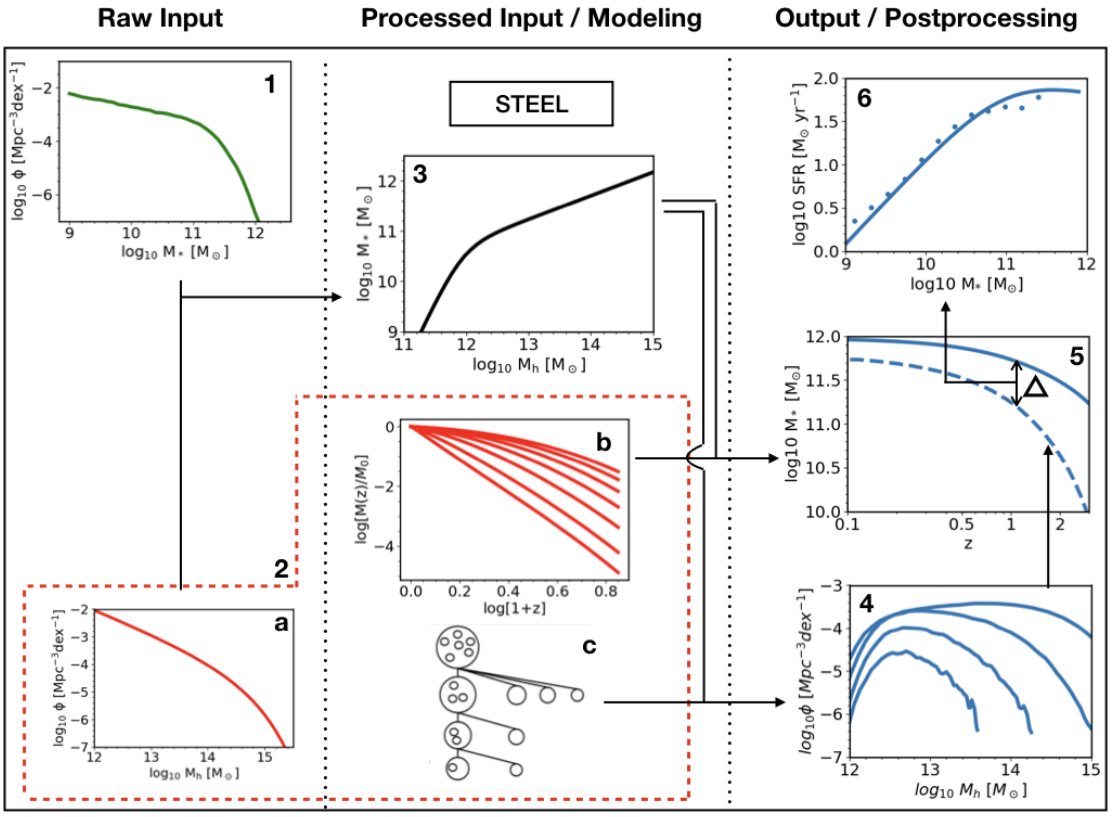


FIGURE 4.4: A cartoon showing the constituent steps of the method to generate star formation rates. In brief, the three columns from left to right are raw inputs, derived inputs/modelling, and output/post-processing. The subplots are: 1. The stellar mass function, 2a. The halo mass function, 2b. Halo mass growth histories, 2c. Accretion histories/Merger tree, 3. The SMHM relation, 4. Group/Cluster satellite richness, 5. Central growth histories/satellite accretion histories, 6. Star formation rate. The star formation rates are derived from the difference between the total growth in stellar mass and that from satellite accretion (panel 5).

The method described above directly links the star formation rate to the accreted mass from satellites. However, in our model satellites grow in mass after infall, we therefore must recalculate the full satellite accretion onto the central galaxies updating their mass using the new star formation rate. Using the updated accretion the star formation rate is recalculated through an iterative process converges very quickly as the updated accretion is found to be nearly identical, as expected from the results of Chapter 3 in which we find the starformation rate to be a second order effect on the masses of satellites. When calculating this difference we also take into account the stellar mass loss rate (MLR) due to stellar recycling using Equations 2.3 & 2.4. The star formation rate - stellar

mass relation derived from this method is fit with a double power law that evolves with redshift given by the following Equation 4.1,

$$
\begin{aligned}
SFR(M_*, z) &= 2N(z)\Big[\Big(\frac{M_*}{M_n(z)}\Big)^{-\alpha(z)} + \Big(\frac{M_*}{M_n(z)}\Big)^{\beta(z)}\Big]^{-1} \\
\log_{10} N(z) &= 10.65 + 0.33z - 0.08z^2 \\
\log_{10} M_n(z) &= 0.69 + 0.71 * z - 0.088z^2 \\
\alpha(z) &= 1.0 - 0.022z + 0.009z^2 \\
\beta(z) &= 1.8 - 1.0 * z - 0.1z^2.
\end{aligned}
\tag{4.1}
$$

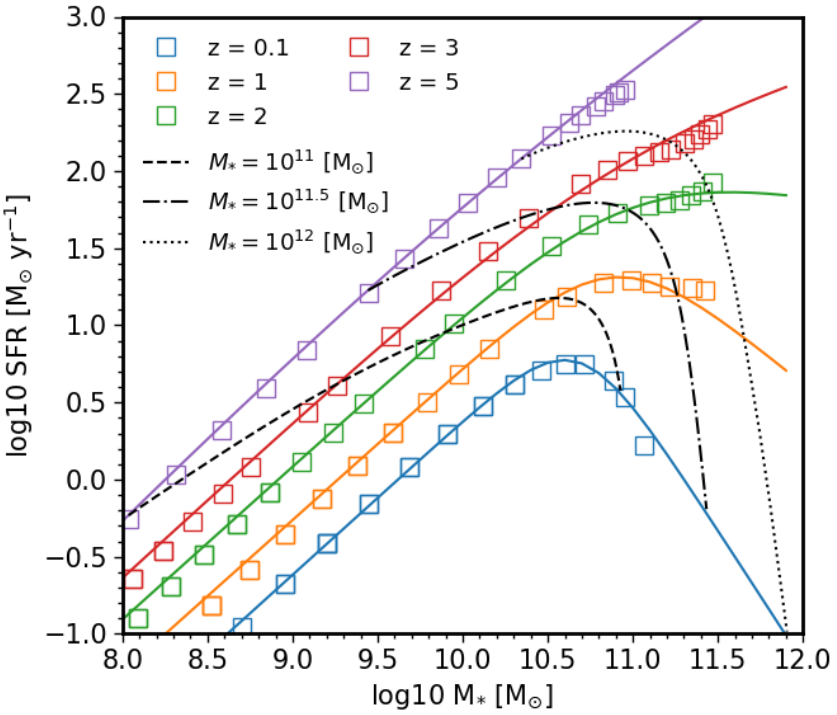


FIGURE 4.5: The star formation rate - stellar mass relation derived from following central galaxy populations along halo mass histories at redshifts $z = 0.1, 1, 2, 3, 5$. The data extracted from the post-processing of STEEL are shown by coloured squares and the double power-law fits are shown as lines in corresponding colours. The three black lines are the evolution on the SFR-M_* plane of the galaxy populations selected at redshift $z = 0.1$ with masses $M_* = 10^{11}, 10^{11.5}, 10^{12}[M_\odot]$ presented in Figure 4.3.

This fit (solid lines) to the computed SFR (open squares) is shown in Figure 4.5. With time (lower redshifts): the normalisation in the mean SFR decreases, the peak of the distribution shifts to lower masses, and the turnover after the peak is steeper. We also show the same three galaxy population tracks from Figure 4.3 as black lines. These tracks show how the galaxy population evolves in SFR with redshift. The population tracks show a gradual increase in SFR and then a turnover before dropping sharply, as they transition to an ex-situ satellite accretion-dominated regime. It is found that

smaller galaxies grow for longer timescales with increasing star formation, whilst larger galaxies start with higher star formation rate and transition to an accretion-dominated phase much earlier in time.

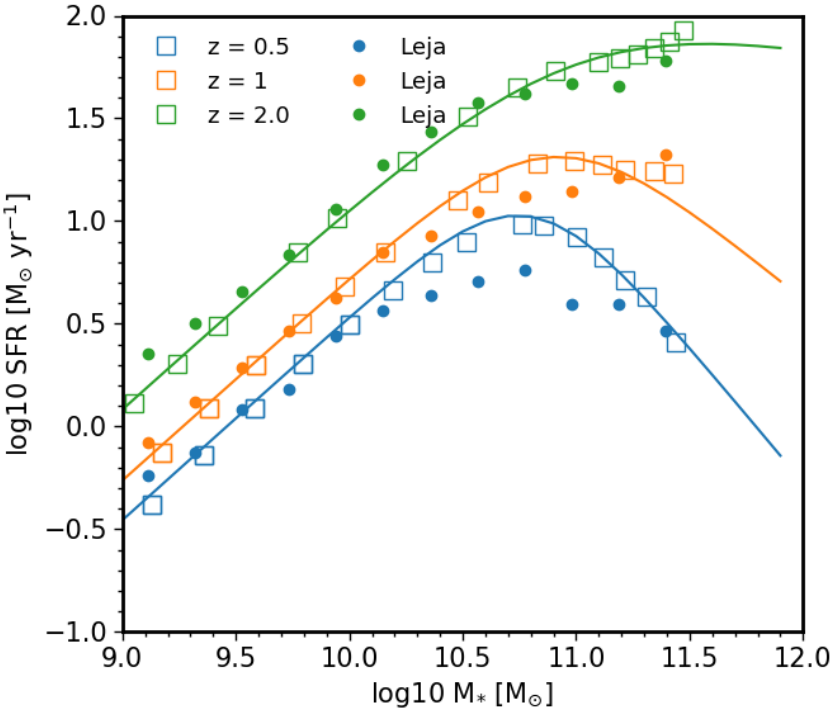


FIGURE 4.6: The Star Formation Rate - stellar mass relationship from Figure 4.5 at redshifts z = 0.5, 1, 2 (blue orange and green respectively, STEEL data are crosses and fits are solid lines) compared with the observed 3D-HST star formation rate from Leja et al. [14] filled circles with corresponding colours denoting corresponding redshifts.

Recent work, where the star formation histories are properly accounted for when measuring star formation rates, has suggested that the previous determinations of star formation rates using UV+IR are 0.1 to 1 dex too high [14] and cannot be reconciled with the growth of the stellar mass function [83, 84]. Our star formation rate is consistent with the results of Leja et al. [14], as reported in Figure 4.6. The excellent match to Leja et. al.'s independent estimates further supports the idea that a more robust method to derive more reliable star formation rates is to follow galaxy assembly along host halo growth histories [see e.g., 17].

In Figure 4.7 we show the cosmic star formation rate evolution predicted by STEEL. The peak at redshift two $z = 2$ is expect as this is widely accepted as the location of the peak of starformation in the Universe. STEEL is a good match to [16] at low redshift and a match to [15] at the high redshift.

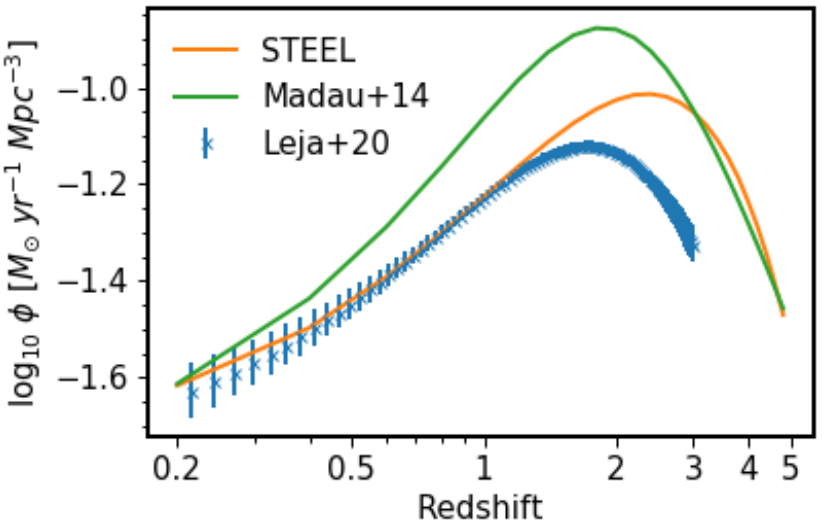


FIGURE 4.7: The cosmic starformation rate density prediction from STEEL (orange line). We include the fit given in Madau and Dickinson [15] (green line) and the data from Leja et al. [16] (blue crosses) for comparison.

4.4.1 Specific Star Formation Rate Distribution

Figure 4.8 shows the specific star formation rate distribution of satellites in three mass ranges, as labelled, chosen to probe transitions found in observational data [159, 163, 164]. The solid blue line and the dashed black lines show the satellite and central sSFR from STEEL, respectively, while the grey histogram shows the satellites from SDSS and the unfilled histogram shows the centrals in SDSS.

STEEL accurately captures the key trends in the distributions, such as bimodality, which is seen in both the central and satellite populations. The central population below $M_* = 10^{10.5}$ [$M_\odot$] is mostly star-forming whereas the satellites show signs of quenching. In the intermediate-mass range, a fraction of the centrals become quenched and the satellites show a strong quenching effect. In the highest mass range, all galaxies show strong quenching features with little star-formation. Whilst still not an exact match to the SDSS distribution, we find that including a redshift dependence in the dynamical quenching provides a better fit than the model used in Paper I. The central sSFR is calculated using the star formation rate presented in Figure 4.5, which uses the PyMorph SMHM relation. Each central mass is assigned a star formation rate with a scatter of 0.2 dex. To account for the fraction of galaxies that are quenched via mergers at each stellar mass we modify the assigned star formation rates by setting a fraction of galaxies equal to the elliptical fraction to have a sSFR of 10^{-12} [yr^{-1}] with a scatter of 0.2 dex and in turn, increase the star formation rate of the remaining galaxies to maintain the same average star formation rate for the population. This approach tests if mergers alone can account for the bimodality found in the central sSFR, the high mass centrals $> 10^{11.3}$

[$M_\odot$], but produces an inadequate fit to the SDSS centrals at masses lower than $10^{10.5}$ [$M_\odot$]. The discrepancies in the location of the star-forming population are likely caused by the imperfect fit to observed SFR as seen in 4.6 and the deficit of quenched galaxies in the lower mass cuts are likely due to causes of quenching that are not merger-related (e.g., AGN feedback).

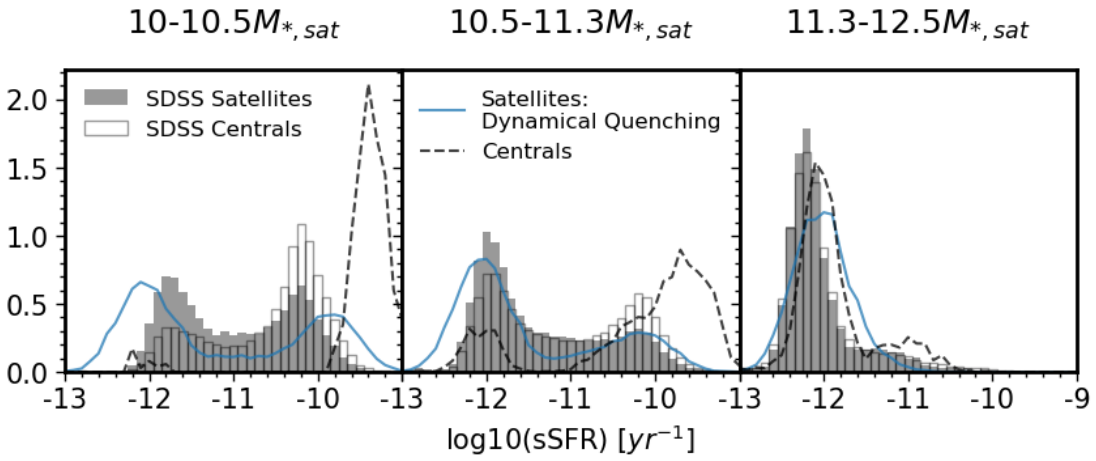


FIGURE 4.8: We show the sSFR of satellites and centrals compared to SDSS in three mass bins selected to mirror proposed breaks in the galaxy main sequence. The SDSS data for satellites and centrals are filled and unfilled histograms respectively. The STEEL result for the satellites is the solid blue line and the post-processed central result is the dashed black line.

4.5 Discussion

In this chapter we found one of the major factors in regulating the in-situ and ex-situ accretion pathways to be the *shape* of the SMHM relation. A shallower low-mass slope causes larger amounts of satellite accretion as smaller haloes, with much higher number density, are initialised with larger satellite galaxies. Similarly to Shankar et al. [82] & Moster et al. [17], we find the high mass slope to undergo only a small amount of evolution with increasing redshift, this implies the growth of central galaxies is directly linked to the steepness of the high mass slope and the growth of the host halo. The flatter the high mass slope of the SMHM relation, the less growth is expected in stellar mass following the assembly of the host dark matter halo. In turn, a weak evolution in the stellar mass content of the central galaxy can be in tension with what is expected from satellite accretion, especially for the most massive galaxies. We discussed that the slope of the high-mass end of the stellar mass function and implied slope of the SMHM relation strongly depend on the choice of light profile, background subtraction, and mass-to-light ratios. However, not all resulting stellar mass functions provide physically self-consistent results in a ΛCDM Universe. Steeper SMHM relations, such as those predicted by PyMorph-based stellar mass functions [122], produce more consistent central and satellite

accretion stellar-mass growths. In addition to models with different SMHM slopes, we also tested models with the dynamical time varied by ±20%, within the range of possible dynamical times predicted in Chapter 3 constrained by satellite richness. This relatively modest alteration has a minor effect on the satellite accretion rate and mass contribution to the central. In this work we find the transitional stellar mass, above which dry mergers progressively become the major contributor to galaxy growth, to be $M_* = 10^{11.1}$, see Figure 4.3. The latter is consistent with previous findings [e.g., 155, 159, 165].
By following the statistical dark matter accretion histories we were able to use the central mass tracks and abundance matching to obtain a growth history for central galaxies. Subtracting from the latter at each time step the cumulative stellar mass from satellite accretion, we created a 'star formation rate' interpreted as the remaining mass required to build the central mass. Our methodology is similar to the continuity approach based on Leja et al. [83] used in Paper I, but with the key difference that here we follow halo growth instead of galaxy number density. The resulting star formation rate for galaxies is notably different to that of Tomczak et al. [166], used in Grylls et al. [167]. At all redshifts, the turnover is notably different, with SFR for masses above the turnover decreasing sharply at low redshift. For masses below the turnover, at $z < 1$ the SFR is lower by 0.3 dex, and at $z > 1$ the SFR is higher by 0.1-0.2 dex. Additionally, the SFR found from this method when combined with morphological quenching arguments reproduces well the bi-modality trends found in sSFR.

4.5.1 Relation to other models

In this Chapter, we show that commonly used stellar mass functions that have been accepted by the community and modellers are inconsistent with ΛCDM cosmological models. The implications of this statement are of interest to wide-reaching areas of galaxy modelling.

- **Distribution of satellites in semi-analytic models.** It has been shown that semi-analytic models struggle to reproduce the galaxy mass function at higher redshifts [e.g. 71]. The results in this chapter present an explanation for this shortcoming of semi-analytic models. A model which attempts to fit a SMF at low redshift, e.g. cmodel, that is not consistent with ΛCDM assembly will experience difficulties in its stellar mass assembly. It is to be expected that the model will attempt to 'fix' these difficulties during the parameter optimisation. High redshift SMF are less well constrained than low redshift SMF it is feasible that a parameter optimisation routine will favour the better constrained SMF. A consistency check between LCDM and the SMFs to be fit or an appreciation of the 'true' systematic error introduced by stellar mass estimation is essential to any holistic model of galaxy formation.

- **Feedback, feedback, feedback.** Galaxy formation models (semi-analytic and hydro-dynamical) rely on a large amount of feedback, i.e. the processes that grow galaxies also lead to suppression of galaxy growth. The two feedback mechanisms that dominate the discussion of stellar-mass growth are super-nova (SN) and active galactic nuclei (AGN), each of these have a large potential energy budget and therefore are excellent ways to reduce the efficiency of star formation. Largely these feedback routines are invoked over two different mass ranges as can be seen on the SMHM relation. SN decrease the efficiency of forming stars in the low mass regime by heating and ejecting gas in/from star-forming regions. In the high mass regime, AGN are thought to eject energy over galaxy/halo scales suppressing starformation globally and quenching galaxies. The knee of the SMHM relation is then the point at which star formation has been most efficient as the sum of these process has least effect [82]. However, the efficiency and coupling strengths with the ISM of the AGN feedback processes are still largely debated [168, e.g.]. If AGN were to be less efficient at quenching, then the result would be an increase in the SMHM relation high mass slope, as we find in our best fit model. It follows that the aggressive feedback found to be required in galaxy modelling is actually a result of trying to suppress star formation to reduce the mass budget from star formation to allow for a better fit to the SMF that are in contention with ΛCDM .

- **Illustris TNG high mass slope.** In Figure 2.8 we show the SMHM relation from the Illustris TNG [23]. It can be seen that the high mass slope is significantly higher than even that produced using the PyMorph SMF. TNG focuses on the reproduction of galaxy structures, galaxy formation, and galaxy clustering, above reproduction of the SMF. The steeper SMHM relation produced by TNG can be expected given that the hierarchical models naturally favour a steep relation.

4.6 Conclusions

In this chapter we build on the success of STEEL in reproducing satellite richness at different cosmic epochs and environments shown in Chapter 3 Figure 3.10. With well constrained high redshift environments we are able to predict reliable galaxy merger rates, using central growth rates calculated using the central mass track of our statistical dark matter accretion histories and abundance matching.

Performing a comparison of the satellite accretion rates and central growth histories. We found that SMHM relations with shallow high-mass slopes create central growth histories that are physically inconsistent with the expected satellite merger rate. We found that steeper SMHM relations at the high mass end, as induced by the latest determinations of the stellar mass functions based on Sersic-Exponential photometry,

are favoured against shallower SMHM relations, based on outdated determinations of the stellar mass function. The total stellar mass growth of a galaxy is mostly due to satellite mergers and/or star formation. A flatter SMHM relation, however, naturally implies, for a given increase in host halo mass, a much weaker growth in the stellar mass of the central galaxy than in the case of a steep SMHM relation. The accretion via satellites could then be substantial enough to overshoot the moderate growth in the central galaxy rendering the model internally physically inconsistent.
By assuming the difference in central growth rate ($\dot{M}$) and satellite accretion rate is attributable to the star formation in the central galaxy, we predict star formation histories and a star formation rate-stellar mass relations. The latter approach is qualitatively similar to a continuity equation [14, 84, e.g.,], but more accurate as it is developed along the accretion tracks of host haloes so better follows galaxy populations. We find our resulting star formation rates to be in excellent agreement with the latest cutting-edge observational measurements by Leja et al. [14], based on multi-parameter Bayesian analysis.

4.6.0.1 Future work

The work presented in this chapter, checking self-consistency between ΛCDM halo assemblies and stellar mass functions, is of the utmost importance to predict robust and self-consistent SMHM relations and to generate reliable mock catalogues for the next generation of extra-galactic surveys such as Euclid and LSST. Furthermore, by refining consistency checking it may be integrated into data analysis pipelines of future extra-galactic surveys, that use for example stellar mass estimation models to create data products such as the SMF. Stellar mass estimation models and cosmological models that do not provide self-consistent growth histories for galaxies can be identified much earlier. To further improve the abilities of an integrated method, additional empirical modelling can be performed such as the star formation presented in this Chapter, the satellite distributions and pair-fractions presented in Chapters 3 & 5, or the morphologies from Chapter 5. Following this method when estimating galaxy features from a survey many assumptions made such as initial mass function, light profile, star-formation rate e.t.c can be checked at point of derivation to avoid the inconsistency issues that have thus far been systemic[2] in galaxy observations.

[2] As discussed:

- SMF and SFR continuity.
- Stellar-mass estimation and ΛCDM assembly
- Stellar-mass estimation and pair fractions

Chapter 5

Galaxy Pairs, Mergers, and Morphologies

"One of the basic rules of the universe is that nothing is perfect. Perfection simply doesn't exist.....Without imperfection, neither you nor I would exist" Stephen Hawking

5.1 Background

ΛCDM cosmology predicts the hierarchical assembly of dark matter haloes. Throughout the history of the Universe, haloes have grown in mass and size via two pathways. Firstly, haloes grow via smooth accretion gradually accreting dark matter from the surrounding environment. The secondary growth mechanism is via the accretion and gradual absorption of smaller haloes, known as subhaloes. After accretion subhaloes survive as the substructure of the central/host halo, gradually losing mass and sinking to the centre of the potential well though dynamical friction. As discussed in previous Chapters, these subhaloes contain satellite galaxies that follow the halo structure resulting in galaxy-galaxy mergers.

Frequent or massive mergers are thought to induce morphological changes in galaxies. Galaxies, after experiencing a massive merger, where the minor galaxy is at least a quarter of the mass[1] of the central galaxy, are thought to lose their disk-like morphology and transform into elliptical galaxies [30, 169]. For this reason, it is important to understand the frequency and nature of mergers between galaxies to achieve a complete and coherent picture of galaxy formation and evolution. Unfortunately, galaxy mergers occur on gigayear timescales and therefore it is not possible to directly observe the rate or consequence of galaxy mergers. The traditional approach to estimate the rate of galaxy

[1] Galaxy merger timescales far eclipse the human lifespan so the merger mass ratio is not an empirical constraint instead inferred from hydrodynamical merger simulations. This has led to the use of the merger mass ratio as a 'free' parameter in simulations which can be used to tune to morphological fractions.

mergers is to count galaxy pairs, and then assign a merging timescale to infer the rate [4, 170–172]. A galaxy pair is usually defined by a mass ratio and a separation. For example, typical choices employed by observers are selecting galaxies that are within 5 to 30 kpc and fit the major merger criteria (1/3 mass ratio). However, the approach of counting pairs is complicated by systematic differences when selecting galaxies. For example, the evolution of the pair fraction appears to change if a selection is made by flux ratio or made by stellar mass ratio [97].

5.2 The systematic effects of stellar mass estimation on galaxy pair fractions

In this Chapter, we show how different SMHM relations generate distinct pair fractions and by association different merger rates. Stellar-mass functions with greater number densities of high-mass galaxies map larger galaxies into smaller haloes, resulting in steeper high-mass slopes for the SMHM relations. In Figure 5.1 we show an illustrative cartoon of how different SMHM relations affect the galaxy mass ratios. For two identical halo pairs, we see that a SMHM relation with a steeper slope causes a substantial difference in the implied stellar mass ratio, when compared to a shallower SMHM relation.

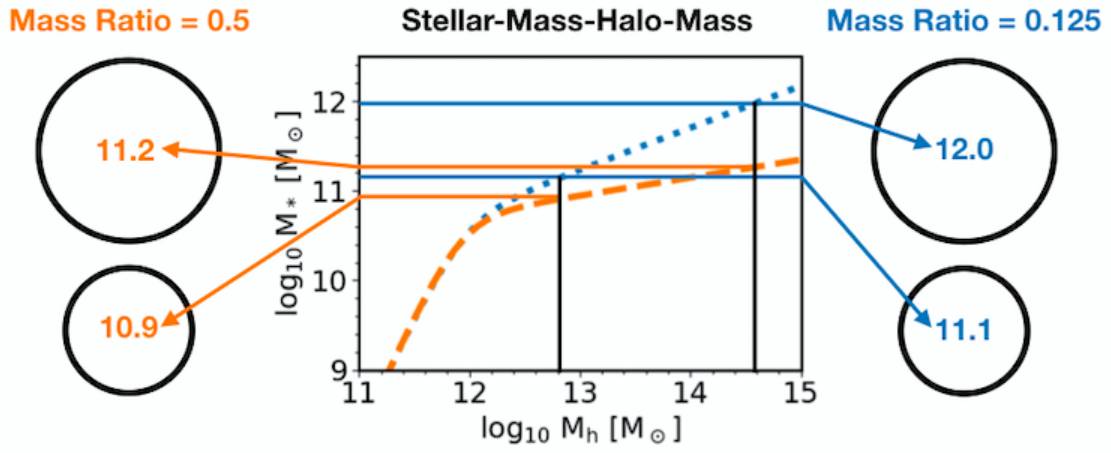


FIGURE 5.1: A cartoon showing how the SMHM relation impacts the stellar mass ratio of galaxies mapped into identical haloes. The steeper SMHM relation creates a smaller stellar mass ratio as the change in halo mass maps to a much larger stellar mass difference.

It follows that given two identical distributions of haloes seeded with galaxies via different SMHM relations the shallower one will seed more galaxy pairs[2].

In Figure 5.2 we show an example of systematic difference expected in the pair fraction when changing the SMHM relation. The left-hand column shows the SMHM relations

[2]The pair fraction is defined here, in relation to major mergers, as the fraction of galaxies of a given mass that have a companion with a mass equal to or greater than a quarter of the primaries mass within 5-30 kpc.

and the right column the pair fractions and their evolution with redshift. In the top row, we compare two high mass slopes, one steep and one shallow, where the slope has been changed at redshift $z = 0.1$. In the bottom row, the slope at redshift $z = 0.1$ is fixed and we compare an evolving and non-evolving high-mass slope.

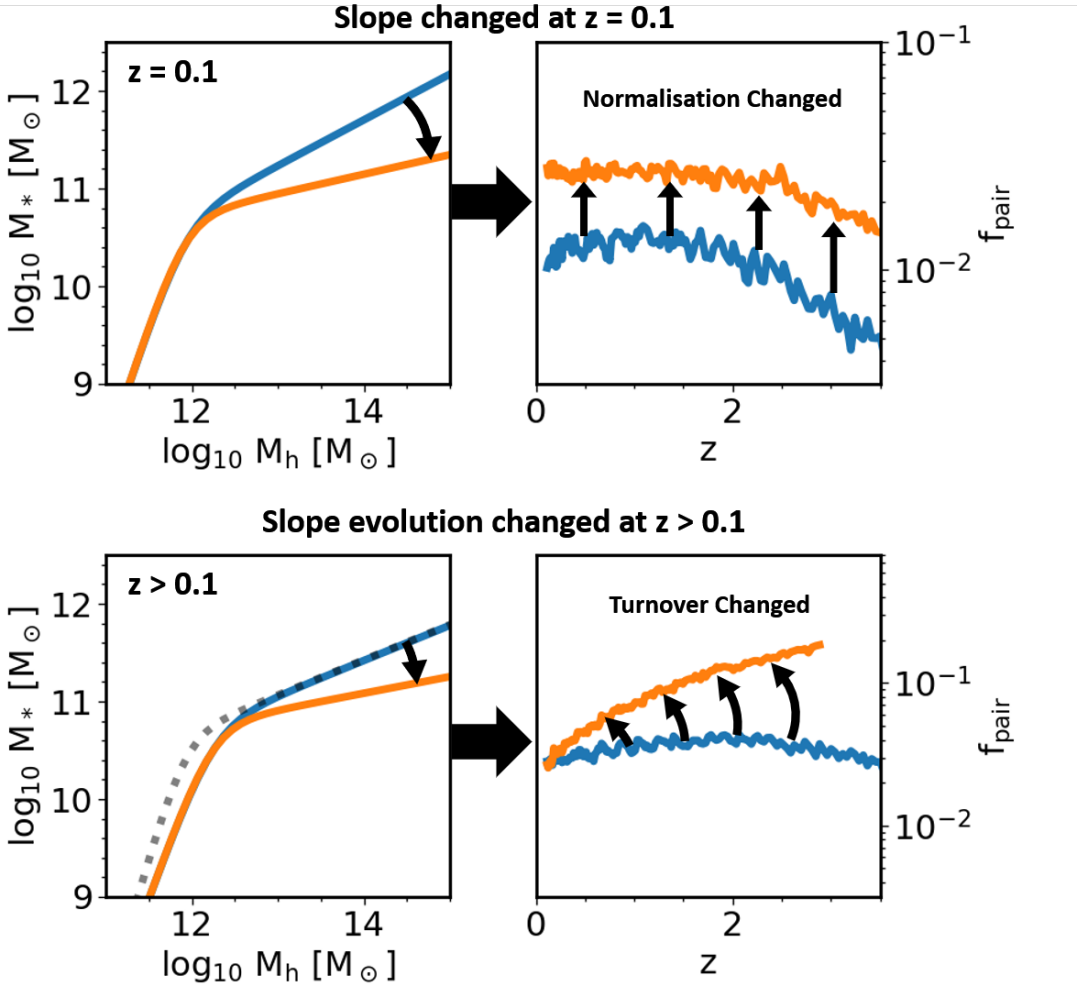


FIGURE 5.2: A cartoon showing how the SMHM relation can impact the pair fraction of galaxies with a mass ratio $> 1/4$. The top row shows how reducing the high mass slope of the SMHM relation increases the number of pairs at all redshifts. The bottom row shows the redshift $z = 0.1$ relation as a grey dotted line alongside two evolving relations at redshift $z = 2$ where one relation is evolved to be shallower (orange lines) compared a relation in remains steeper (blue lines). For the SMHM relation that evolves to be shallower, the pair fractions are found to increase. In each case, the reason for the increase can be explained via Figure 5.1, which shows that shallower SMHM relations produce a greater number of massive pairs.

The suppression of the high-mass slope increases the number of pairs created and the normalisation of the pair fraction increases. In the bottom row, we show the effects of having a slope that flattens at higher redshift. We show the redshift $z = 0.1$ relation in grey and the relations with the unchanged and changed slopes in blue and orange, respectively. The main effect of varying the evolution of the high-mass slope in the

SMHM relation is to change the behaviour of the pair fraction with redshift. A steeper slope leads to a decreasing pair fraction, whereas a slope that gets shallower with time results in an increasing pair fraction.

The behaviours reported in Figure 5.2 are a direct consequence of the trends sketched in Figure 5.1, where shallower slopes give higher fractions. Furthermore, from Figure 5.2 (and from Figure 5.3), it can be concluded that almost any pair fraction difference could be produced by appropriately altering the input SMHM relation. It is relevant to stress here that relatively minor changes in the stellar mass function can cause qualitative differences in the SMHM relation and, by extension, in the shape and normalisation of pair fractions at any cosmic epoch.

The ability to systematically change the pair fraction due to stellar mass derivation calls to question the discrepancies found in pair fraction results [e.g. 97]. Systematic differences in the stellar mass measurements from different groups, will naturally yield different SMFs and thus, by extension, different SMHMs and pair fractions. STEEL as a flexible and lightweight model can shed light on the extent of the systematics in the predicted pair fractions implied by different stellar mass estimates and SMFs, by running the same underlying cosmological model with the input SMHM relations corresponding to each distinct SMF.

Calculation of the pair fraction in STEEL requires an estimate of the distance between the central galaxy and the satellite galaxy. In our analytic statistical DM accretion histories, we lack 'physical' separations, we instead assign each subhalo bin an average distance to the central galaxy. The subhaloes start at the viral radius of the central halo, the distance to the centre then reduces directly proportionally to the amount of dynamical time elapsed [31], e.g. once half the dynamical time has elapsed the distance will be equal to half the starting distance.

In this Chapter, we investigate the impact on the distribution of satellites around central galaxies when using different SMHM relations. As this distribution is primarily dominated by the halo substructure, it is essential to make sure our selection criteria for galaxies always returns the same halo population. In a simulation with known haloes they can be directly selected. However, to better match observations, where haloes are not known, the selection must be performed on an observable property. In STEEL, the stellar masses are a function halo mass and therefore cannot be used. Instead, a constant number density selection can be made, selecting objects by abundance which theoretically will return the same halo population as we assume the halo structure to be a fixed quantity. A similar technique was employed by, van Dokkum et al. [e.g. 173], Huertas-Company et al. [e.g. 174], Leja et al. [e.g. 175], Mundy et al. [e.g. 176] to trace the evolution of galaxy populations. Galaxies at high redshift with a given comoving number density are assumed to be the progenitors of galaxy populations observed at later epochs with the same abundance. In this work we use a central stellar mass selection from

TABLE 5.1: The adjustments to the SMHM relation used in Figure 5.3.

	PyMorph	$X_{0.1,alt}$	$X_{z,+}$	$X_{z,-}$
M	11.92	-0.25	-	-
M_z	0.58	-	+0.1	-0.1
N	0.032	+0.04	-	-
N_z	-0.014	-	+0.007	-0.007
β	1.64	-0.3	-	-
β_z	-0.69	-	+0.3	-0.3
γ	0.53	+0.06	-	-
γ_z	-0.03	-	+0.2	-0.2

the PyMorph stellar-mass estimation of $M_* = 10^{11} - 10^{11.6} M_\odot$ (or $10^{9.5} - 10^{10.1} M_\odot$ when considering the low-mass slope controlled by β). For any given input SMHM relation we then select the galaxies that have a number density equal to the one corresponding to the aforementioned mass cut in the reference stellar mass function. An example of this selection can be seen in Figure 5.3. The shaded horizontal band shows the stellar masses for each SMHM relation with equal number density.

The SMHM relation is defined by 4(+4) parameters. We build a toy model where each of the parameters (M, N, β, γ), and their evolutionary factors (M_z, N_z, β_z, γ_z), are adjusted in turn to explore the effect on the galaxy pair fractions. Table 5.1 details the change made to the SMHM relation for each parameter.

Figure 5.3 shows each of the SMHM relations in the outer four panels. The reference SMHM relation derived from the PyMorph photometry is depicted in blue at redshifts $z = 0.1$ (dotted line) and $z = 2$ (dashed line) in each panel. The modified redshift $z = 0.1$ relation is then shown in orange, and the increased and decreased (dashed red and green) evolution relations are plotted at redshift $z = 2$. The inner four panels show the results on the pair fraction and follow the same colour convention.

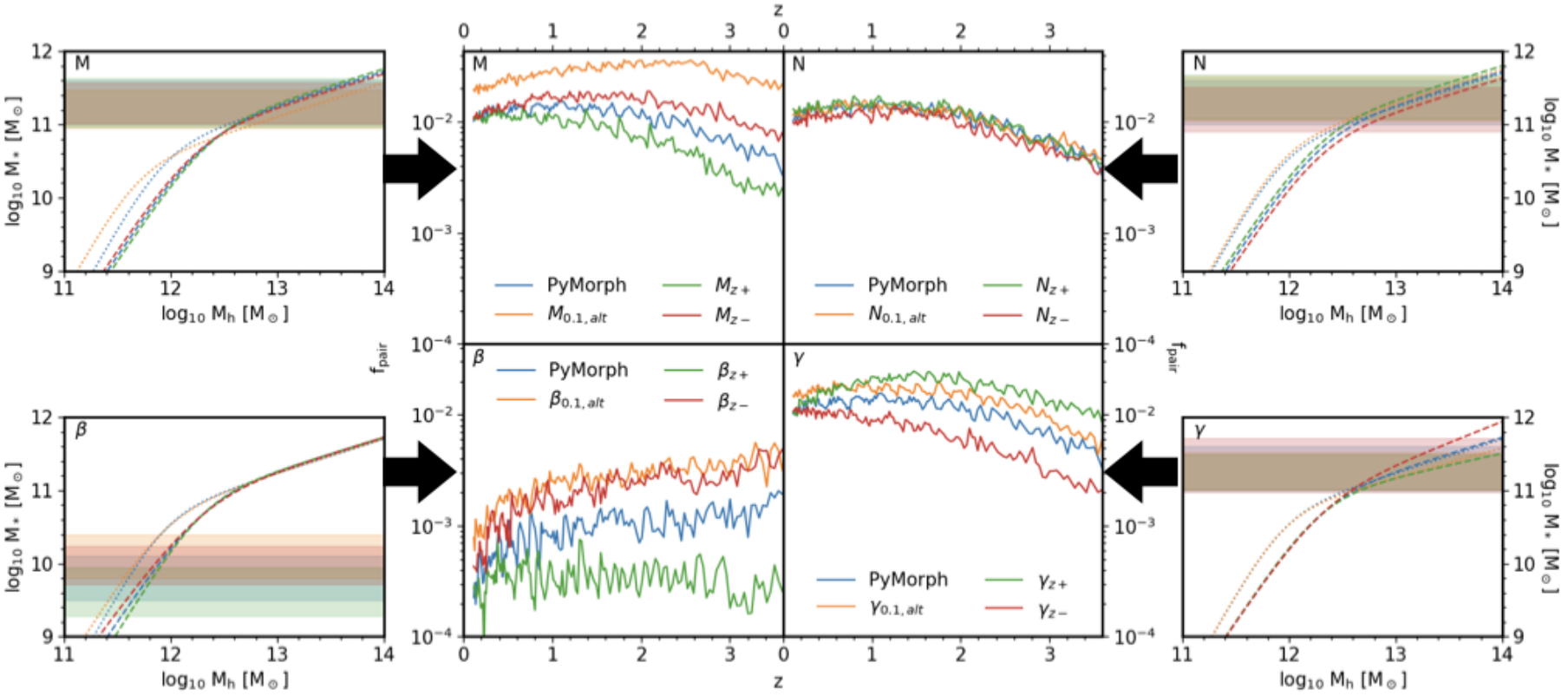


FIGURE 5.3: Each of the panel pairs (M, N, β, γ) shows the input SMHM relation in the outer plot and the modelled pair fraction evolution in the centre plot. Each pair investigates adjustments to the given parameter of the SMHM relation (M, N, β, γ). Each pair shows the reference SMHM relation 'G18' in blue, the relation adjusted at redshift $z = 0.1$ keeping the same SMHM relation evolution parameters in yellow. The red and green lines respectively have the evolution parameter altered such that the evolution parameter is increased or decreased with respect to PyMorph. In the outer (SMHM relation) plots, dotted lines are $z = 0.1$ relations and dashed lines are $z = 2$ relations. The PyMorph relation is shown at both epochs for comparison. Finally, the shaded bands in the outer plots show the consistent number density selections used in the centre plots.

When changing M, the knee parameter, a large increase in the pair fraction is found from a lower knee: The shallower high mass slope is extended, therefore more haloes are seeded in the mass range for pair selection. We see the same effect at high redshift, the lower value of M creates a higher pair fraction. The normalisation parameter, N, creates little change in the pair fraction as expected because the mass ratios are largely unaffected. The low mass slope parameter, β, affects the seeding of smaller galaxies hence a lower mass range is used for the consistent number density cut. Due to the steepness of the low-mass slope, the fraction of pairs found is lower in this mass cut. Finally, when the high-mass slope parameter, γ, is altered, more pairs are found at high and low redshift when the slope is shallower. This is again attributed to more galaxies seeded within the mass ratio range.

5.2.1 Comparison to simulation and observational results

The observed pair fraction is known to have discrepancies based on the galaxy property used to calculate the ratio. In Man et al. [97] it is shown that selecting pairs by flux ratio or stellar mass creates differences in the pair fraction evolution. In Figures 5.1 & 5.2 we show the predicted effect of the determination of stellar mass on the SMHM relation and the systematic propagation of these changes into the pair fraction. Through the use of a toy model, in Figure 5.3 we show how isolated perturbations to the eight SMHM relation parameters propagate into the galaxy pair fraction. From this analysis, we conclude that any measurement of the pair fraction is strictly related to the underlying assumptions made during the processing of observations, that affect the measurement of stellar mass. Our findings thus imply that meaningful comparisons between different pair fraction measurements can only be undertaken under identical stellar-mass derivation assumptions, or where this is not the case the influence of any differences are accounted for. In this section we fit, by making use of STEEL, observed pair fractions using small changes to the SMHM relation. We anticipate this modelling can be used to provide corrections to pair fraction results to allow for more meaningful comparisons.

5.2.1.1 Simulation: Illiustris TNG

In Figure 5.4 we show the simulated galaxy pair fractions for galaxies in the mass range $M_* = 10^{10} M_\odot$ to $10^{10.6} M_\odot$. The pair fraction is shown for two different SMHM relations used as inputs to STEEL. In blue we show the PyMorph (Sèrsic-Exponential) input used as the reference in Figure 5.3, in orange the input SMHM relation is calibrated to match outputs of the Illustris TNG simulation. In the right-hand panel, we see that pair fraction predicted by STEEL using in input the SMHM relation calibrated on TNG is in good agreement with the pair fraction extracted directly from the Illustris TNG simulation. This is a valid test that lends further support on the ability of STEEL to generate robust

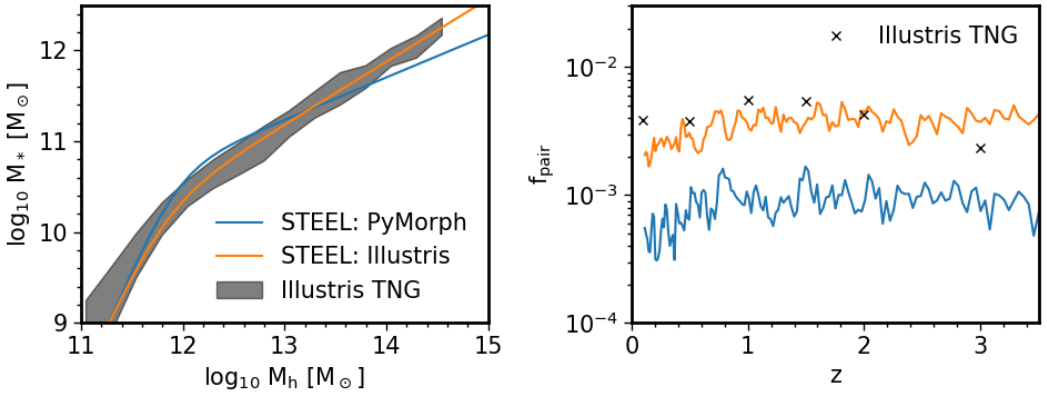


FIGURE 5.4: Left: Two SMHM relations are shown from STEEL using parameters designed to reproduce the SMHM relation found in the Illustris TNG simulation (Orange line) and the PyMorph (Sèrsic-Exponential) fit parameters (Blue line). The shaded region is the output from the Illustris TNG simulation. Right: The pair fractions predicted by STEEL for galaxies in the mass range $M_* = 10^{10} M_\odot$ to $10^{10.6} M_\odot$ from the two input SMHM relations plotted in the left panel (same colour coding). The black crosses indicate pair fractions directly extracted from the Illustris TNG simulation.

pair fractions for a given input SMHM relation. The pair fraction predicted using the PyMorph input is 0.5 dex lower. This is expected as in the mass range we are considering the Illustris TNG simulation SMHM relation is shallower and more pairs are therefore created in a greater mass range of halo mergers.

5.2.1.2 Observation: Mundy et al. [4]

To further investigate the variations of pair fraction when varying the input stellar mass estimates and therefore the SMF, we compare the predicted pair fraction evolution using two SMHM relations. Specifically we use the SMF presented in Section 2.4.2, PyMorph and cmodel. PyMorph has a substantially greater number of high mass galaxies and thereby produces a steeper SMHM relation. In Figure 5.5 the left-hand panel shows each SMHM relation at redshift $z = 0.1$ and $z = 2.5$. Following the systematic investigation in Figure 5.3, we attribute the 0.1 dex difference in pair fraction to the difference in high mass slope between PyMorph and cmodel. The best-fit relation from Mundy et al. [4], shown as black crosses, rises at earlier cosmic epochs rather than falling; as seen from PyMorph and cmodel. We saw in Figure 5.3 that a SMHM relation with a high-mass slope decreasing with redshift will generate a pair fraction evolving similarly with redshift.

In an attempt to reproduce the Mundy et al. [4] pair fraction evolution with redshift we begin using the cmodel SMHM relation which gives the closest match to Mundy et al. [4] in pair fraction at low redshift. Following the analysis of Figure 5.3, where higher γ_z increases the pair fraction at high redshift, we alter the parameter from 0.0 to 0.5 in steps of 0.1. In Figure 5.6 the left panel shows the SMHM relation at redshift $z = 0.1$

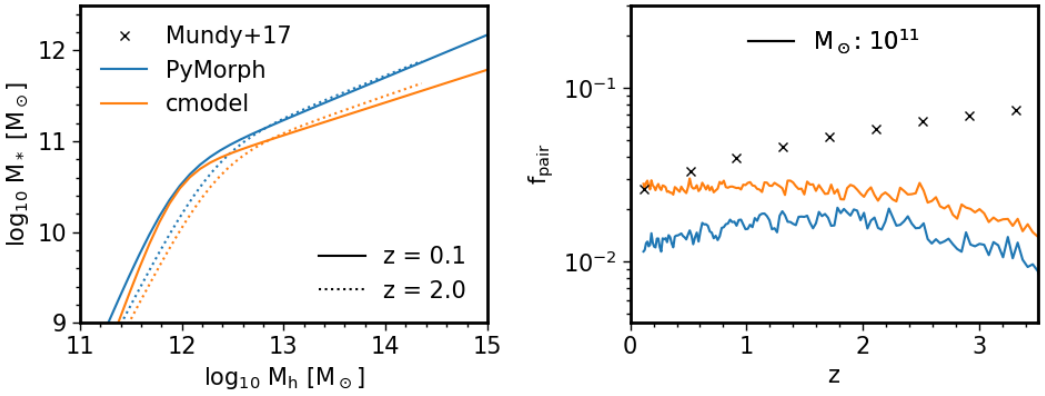


FIGURE 5.5: Left: Stellar-mass-halo-mass relations derived from PyMorph (blue) and cmodel (orange) at redshifts $z = 0.1$ (solid lines) and $z = 2.0$ (dotted lines). Right: The pair fraction evolution for galaxies using both SMHM relations. The black crosses show the corresponding best fit for the $> 10^{11} M_{\odot}$ mass cut from Mundy et al. [4].

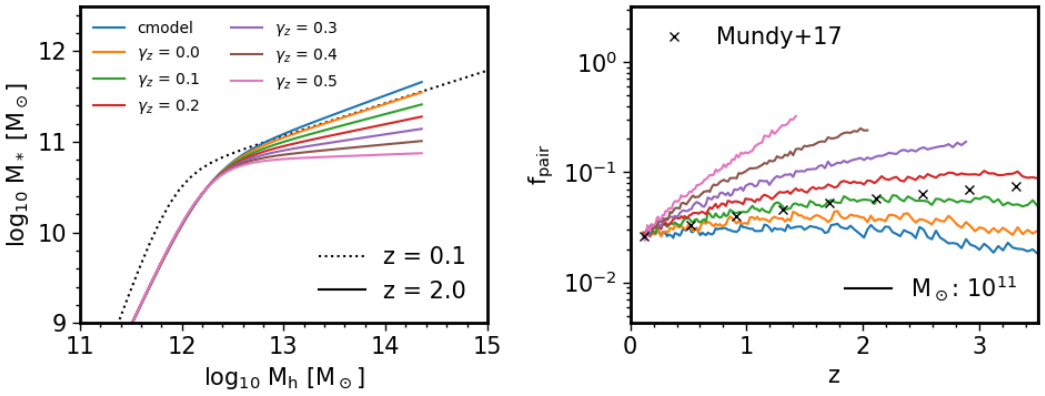


FIGURE 5.6: Left: Stellar-mass-halo-mass relations derived from cmodel (black) at redshift $z = 0.1$ (dotted lines) and at $z = 2.0$ with altered high mass-slope evolution parameter (coloured lines). Right: pair fractions corresponding to the SMHM relations plotted in the left panel (same colour coding). The black crosses show the best fits for the $> 10^{11} M_{\odot}$ mass cut from Mundy et al. [4].

as a black dotted line, while the coloured lines show the relation at redshift $z = 2$ with different γ_z parameters. The right panel shows the impact of this evolution on the pair fraction. As predicted higher γ_z increases the pair fraction with redshift and a value of above 0.1 removes the turnover. Comparing to Mundy et al. [4] we see a value of γ_z between 0.1 and 0.2 best reproduces the rise in pair fraction.

5.3 Galaxy Morphologies Resulting from Galaxy Mergers

Mergers are thought to be one of the drivers for morphological transformation, size growth and other structural and dynamical changes in galaxies [98, 177–180]. A number of analytically-based cosmological models have generally assumed that major mergers

in particular, with a mass ratio of at least $M_{sat}/M_{cen} > 0.25$ (to 0.3), are effective in destroying disks and in forming ellipticals [181–183]. However, galaxy mergers occur on megayear or even gigayear timescales, therefore, observations of galaxy mergers are essentially static snapshots of the processes. Most of our knowledge of galaxy mergers comes from analysis of simulations of merging systems [e..g. 32, 65, 99, 178, 184–188]. Despite the wealth of simulations of galaxy mergers, the fulcrum of our understanding of how mergers affect the galaxy population lies with the estimation of the merger rate and significance of mergers [32, 185]. Using the merger rates derived in Chapter 4 we implement a post-processing method to ascertain the effects of traditional mergers models when applied to true, statistical, populations in STEEL.

5.3.1 Implementing discrete processes in a statistical model.

One of the primary challenges in the development of STEEL is trying to implement discrete processes in a statistical fashion. Mergers between two galaxies are implemented stochastically using the following method.

- At each time-step, for each central halo mass track, the mass bins of previously accreted subhalo/satellite galaxy mass functions that have reached the end of their dynamical time, are summed to produce the merging satellite stellar mass function.
- Each central halo mass track is assigned a central stellar mass at each epoch $M_{*,cent}(M_h, z)$ (Figure 2.10). By integrating the merging satellite stellar mass function in the range $[M_{*,cent}(M_h, z) \cdot \mu,\ M_{*,cent}(M_h, z)]$ the probability that a given central has undergone a major merger at each epoch is retrieved, where μ is the major merger mass ratio.
- In Figure 5.7 an illustration of the way galaxy morphologies are updated is given.
  - In the leftmost panel, all galaxies are spirals or a disk-like morphology (blue bar). The probability of a galaxy having a major merger is shown as a black line.
  - Following the arrow to the middle panel the fraction of galaxies that had a major merger are assigned as ellipticals. The mass track at this epoch now has this spiral to elliptical ratio.
  - At the next time step, the fraction of galaxies undergoing a major merger is calculated. However, this fraction is split between galaxies that have previously had a major merger and those that have not.
  - Only the galaxies that were in the spiral group in this time step contribute to the increasing elliptical fraction.

– Repeating this procedure at each time-step eventually all galaxies may have the potential become elliptical but the rate at which this happens progressively slows in time due to the decreasing spiral population.

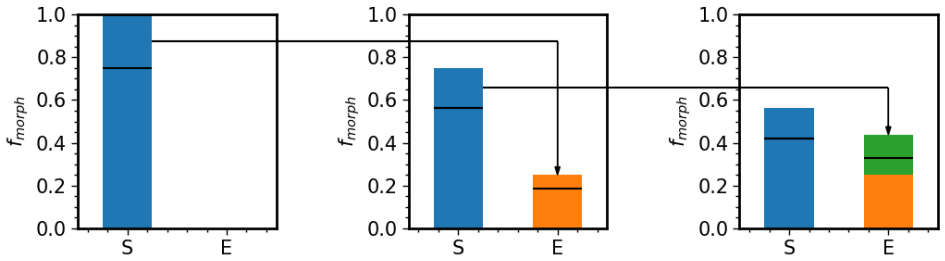


FIGURE 5.7: A cartoon to show the way we assign morphologies statistically in STEEL. Each step has the same fraction of major mergers but the number of ellipticals created reduces as some major mergers occur on the elliptical fraction. The fraction of galaxies in each population experiencing a major merger is displayed as a horizontal black line.

Applying the above process to galaxies in STEEL from redshift $z = 3$, we build the morphological fraction of galaxies at different masses using a major mass ratio of $\mu = 0.25$, at redshifts $z = 0.1, 0.65, 1.75$. Figure 5.8 shows the probability/fraction of central galaxies that have elliptical morphologies, while the black triangles show the T-Type-selected elliptical fraction from the SDSS catalogue at redshift $z = 0.1$. We find that applying this simple recipe to the merging number densities from STEEL creates a good match to the elliptical fraction in the local universe.

5.4 Lenticular Galaxy Formation

Lenticular galaxies are categorised by their unusual appearance. In the famous 'tuning fork' diagram derived by Hubble [25], the lenticular galaxies are found where the two spiral 'forks' meet the elliptical 'handle'. Lenticulars similarly to spirals have an extended disk but do not boast the 'bars' or 'arms' that decorate the disks of spiral galaxies nor the reservoirs of cold gas suitable for star formation [189]; similar to ellipticals lenticulars have a velocity dispersion-dominated bulge at their centre.

Given that are morphologically and dynamically positioned in between spirals and massive ellipticals, it is reasonable to assume that they may have arisen on the transformation pathway between the two populations. The latter hypothesis is further corroborated by the observational evidence that lenticulars [190], and the fraction of lenticulars increases with decreasing redshift similarly to that of ellipticals [191].

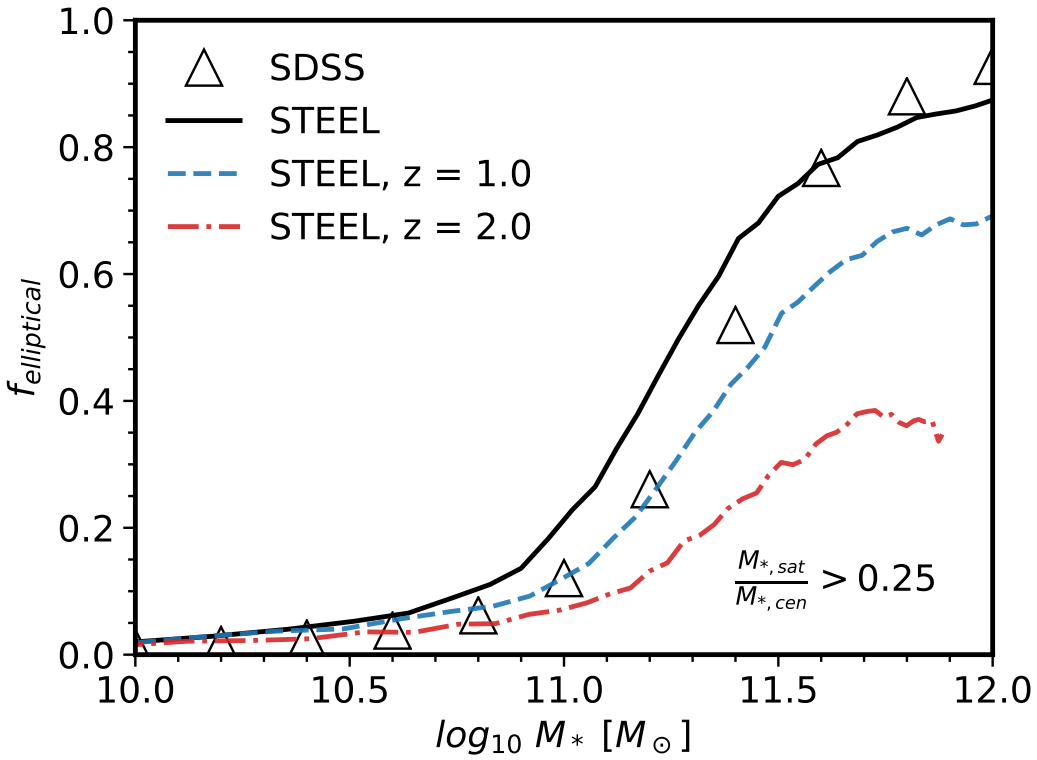


FIGURE 5.8: Fraction of ellipticals as a function of stellar mass predicted by STEEL at three different redshifts, as labelled. The triangles are the T-Type selected elliptical fraction from SDSS at redshift $z = 0.1$.

5.4.1 Lenticular formation in empirically constrained environments.

The work presented in this subsection was carried out by SP under the supervision of PJG and FS. The direction of the project was heavily influenced by PJG and builds upon the successes found in predicting morphologies in STEEL .

Given the successes in ratifying the simple merger model of elliptical formation within STEEL, we extend the predictions of STEEL to lenticular galaxies. We employ an additional assumption (and related parameter) in the model, and compare with SDSS data.

5.4.1.1 Cook et al. [5] Model

We started our investigation by implementing the lenticular formation model by Cook et al. [5]. This model divides halo growth into two distinct phases:

- A fast collapse and intense merging phase at high redshift.
- A slow accretion phase at low redshift after a transition epoch z_t.

Following this two-phase halo growth, the central galaxy first forms a bulge during the fast collapse phase, and then gradually forms a stellar disk around it in the second

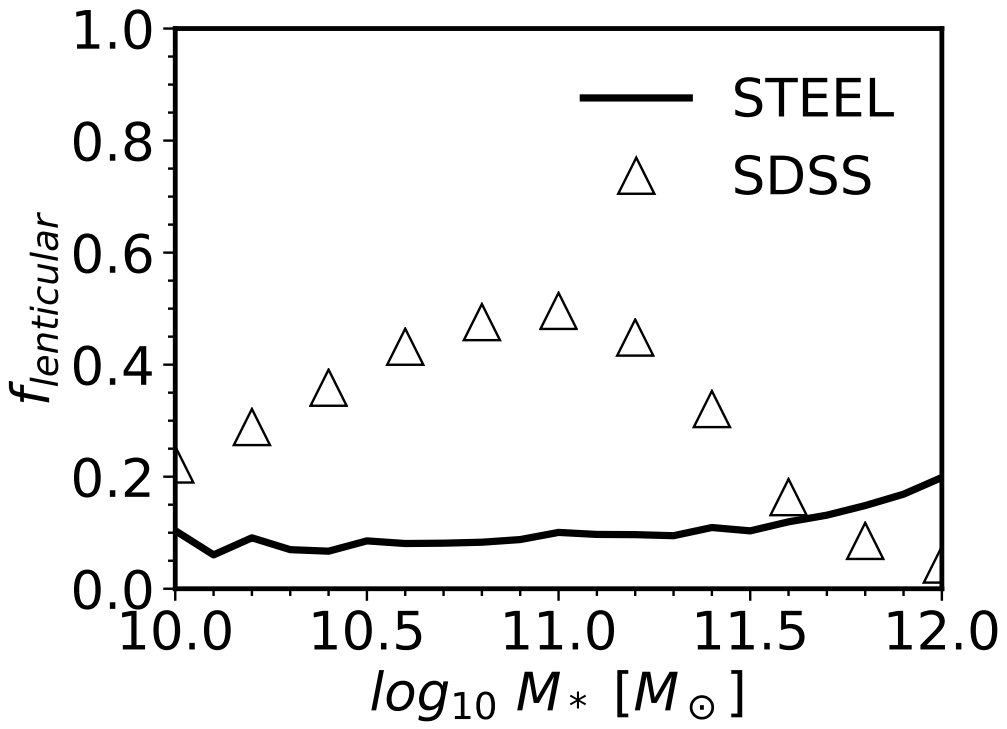


FIGURE 5.9: The solid line marks the lenticular galaxy fractions generated using a simplified version of the lenticular formation method described in [5]. The model assumes lenticular galaxies are those that have formed a significant fraction of their stellar mass during a particular epoch. However, even after significant optimisation of the mass threshold and epoch definitions, it is never possible to reproduce the SDSS data (open triangles) within STEEL using this model.

phase. The model for building spheroids in this semi-analytic model is the same as that implemented in Granato et al. [150].

Translating semi-analytical techniques into a statistical-semi-empirical model by the nature of the two models results in a loss of complexity. In order to maintain the core prediction of the Cook et al. [5] model, we focus on the two-phase formation aspect as follows.

Galaxies that have grown a significant fraction of their mass (e.g. 70%) between two epochs are assigned as lenticulars. These galaxies can then undergo further evolution by being transformed into ellipticals as described by the merger model from the previous section. Despite exploring a large parameter space in the mass fraction and redshift thresholds, we still find that within the framework given by STEEL, the Cook et al. [5] model can not reproduce the SDSS measured lenticular fractions, as illustrated in Figure 5.9.

5.4.1.2 Minimal assumption model

Given the unsuccessful implementation of the Cook et al. [5] model, we revisit other proposed models to find out which (or a set of) have the potential to generate the observed lenticular fraction. Considering the nature of lenticulars in observations, which increase their fraction with decreasing redshift in a similar fashion to ellipticals [174, 192] the first model we implement extends that of the elliptical merger model. An additional merger mass threshold (μ) is implemented at $\mu = M_{*,sat}/M_{*,cen} = 0.05$. Central galaxies that experience a merger between this mass threshold and the major merger threshold are assigned to be lenticulars, while the central galaxies with mu>0.25 are still converted into ellipticals. Lenticular galaxies that subsequently experience mergers above the major mass ratio are then reassigned as ellipticals. This essentially adds a central 3rd column to Figure 5.7. As can be seen in the first panel of Figure 5.10 this simplistic model over produces lenticulars in the high mass end.

To reduce the lenticular production at high masses, following similar ideas to the gas dissipation models utilised in Hopkins et al. [193], we test a hard gas threshold. Physically this gas threshold can be interpreted as a damping force, the gas dissipates the energy of the merger into heating the gas reservoir limiting the disruption to the galaxy structures. Initially, only galaxies with a gas fraction above the threshold can be converted into lenticulars. As shown in the middle panel this removes all lenticulars from the high mass end where galaxies tend to be gas-poor. Finally, to improve the fit to the data we implement a 'soft' gas fraction, according to which galaxies below a given mass fraction become statistically less likely to become lenticulars after experiencing a major merger. We include the soft gas threshold to reflect the varying efficiency of gas at different fractions dissipating the energy. This combination of models well fits the high-mass lenticulars as can be seen in the rightmost panel of Figure 5.10.

Lower mass galaxies reside in less rich environments and grow more via secular (in-situ) processes. It is therefore expected that the merger/dissipation model will not significantly contribute to the formation of lenticulars at these masses. To build lenticulars at these masses we implement a disk instability model following the baryonic inflow rates given by Bournaud et al. [194]. During their continuous growth in stellar mass along cosmic time, we allow galaxies to build their central bulges through gas transport from the disk using Equation 5.1 the analytic approximation of Bournaud et al. [194] extracted from high-resolution hydrodynamic simulations,

$$\dot{M}_{inflow} = 25 log_{10}\Big[\frac{M_{*,disk}}{10^{11} M_{\odot}}\Big]\Big(\frac{1+z}{3}\Big)^{1.5}. \tag{5.1}$$

The mass of the bulge is then given by

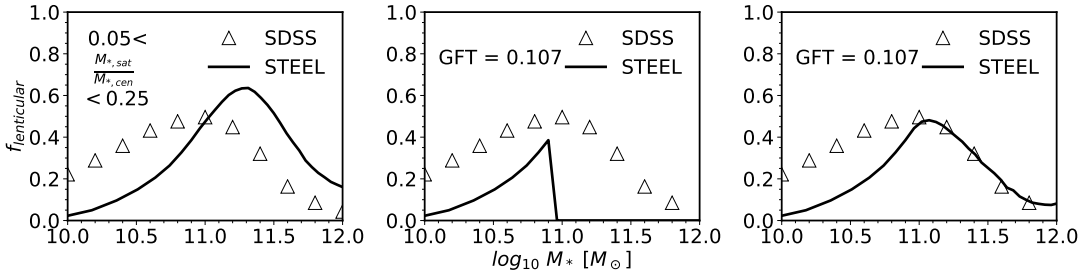


FIGURE 5.10: Three plots from left to right show lenticular fractions generated using the merger model with different parameters and variants. The first (leftmost) panel shows the fraction generated by summing the number densities of merging galaxies with mass ratio in the range $0.05 < \mu < 0.25$. The middle panel shows the effect of adding a hard gas ratio cutoff to the merger model, galaxies that are below a gas fraction threshold (GFT) do not form lenticulars. Finally, changing the model to a 'soft' gas fraction threshold where the efficiency of lenticular formation is reduced above the GFT.

$$M_{bulge} = \sum_{z} \dot{M}_{inflow} \times \delta t(z). \tag{5.2}$$

In each mass bin a fraction of galaxies proportional to the mass ratio of the bulge to the total stellar mass, $A \times M_{bulge}/M_{*,total}$, are added to the lenticular population. The resulting fit to the population is shown in Figure 5.11.

We find the combination of a merger and inflow model is able to adequately recreate the lenticular fractions. This two-channel model is potentially consistent with the distinct morphological types, barred and un-barred, found within the lenticular population [195, 196].

In summary, to recreate the lenticular population we find it necessary to add three assumptions/parameters in addition to the major merger mass ratio used to create the ellipticals. The first is the 'minor' mass ratio limit that defines the minimum mass ratio at which lenticulars are formed via mergers, the second is the 'soft' gas mass threshold that damps the creation of lenticulars, and finally, we need to allow for an additional channel of bulge growth via secular evolution, which, at variance with the merger channel, is particularly effective at lower stellar masses.

5.5 Discussion

In this chapter, we have shown two applications of STEEL. Firstly, we proved the ability of STEEL to unravel the effects of systematics in stellar mass estimates in varying the implied galactic pair fractions. Secondly, we show an example of how constrained merger histories can be used to test the efficiency of morphological evolution models reproduce the observed morphological fractions.

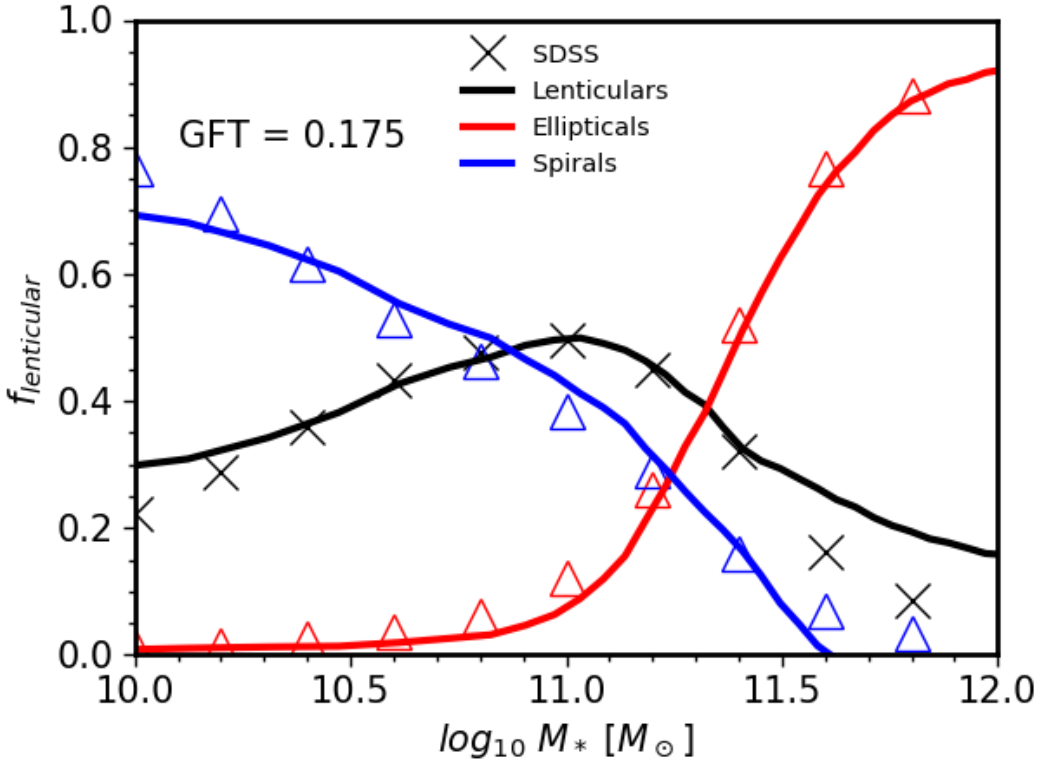


FIGURE 5.11: Fractions of different galactic morphological classes as a function of stellar mass. Ellipticals are generated using a major merger model, lenticulars using the combined merger (with soft gas fraction) and baryonic inflow models and, spirals are the remaining population. Models are compared to data from SDSS (symbols, as labelled).

5.5.1 Pair Fractions

The primary goal of the pair fraction analysis is to show the propagation of systematics in galaxy modelling. Specifically, we used the SMHM relationship to connect assumptions used when estimating observed stellar masses to systematics in galaxy pair fractions in the context of a ΛCDM Universe. In this chapter, we adopted as a reference two observed stellar mass functions from SDSS-DR7 observations that use a de Vaucoulers and a Sèrsic + Exponential fit to determine stellar masses, which in turn create stellar mass functions with notably different number densities at the high mass end. Each stellar mass function then generates, through abundance matching, a different SMHM relationship. The Sèrsic + Exponential mass function generates a steeper high-mass slope in the SMHM relationship at any epoch. In addition to the SMHM relationships from the observed data, we used a relationship fitted to match the outputs of the Illistris simulation. Furthermore, we also considered a toy model SMHM relation individually perturbing each input parameter to transparently probe the impact of the input SMHM relation on the pair fractions. In each case we find that small changes introduced into the

SMHM relationship can have significant effects on the expected pair fractions, as shown in Figures 5.1,5.2, & 5.3. This suggests that in the context of a ΛCDM Universe tensions in previous observational studies could, in large part, be traced back to systematics in stellar mass estimates.

In Mundy et al. [4] the $M_{*,cen} > 10^{10} M_{\odot}$ pair fraction is not significantly different from the $M_{*,cen} > 10^{11} M_{\odot}$ pair fraction shown in Figure 5.5. In Figure 5.3 we find the pair fraction drops significantly when a mass selection is taken below the SMHM relation knee. As this drop is not found by Mundy et al. [4] we interpret that their pair fraction measurement is not consistent with a break in the SMHM relation between $10^{10} M_{*,cen}$ and $10^{11} M_{*,cen}$.

Man et al. [97] noticed that the choice between luminosity-selected and stellar-mass selected pairs affected the pair fraction evolution. In this work, we provided a clear framework to properly interpret how to input choices create systematic effects in the observed pair fraction and its evolution. Furthermore, it is a common approach to infer the assembly history of galaxies by converting the pair fractions into merger rates by assigning timescales to galaxy pairs [4, 170, 171].

In this Chapter, we connected the shape and evolution of the SMHM relationship to the evolution of the pair fractions. We proposed it is therefore possible to use the pair fraction as an additional constraint to the SMHM relationship. This is a natural extension of conditional abundance matching or extended SHAM (subhalo abundance matching) models [197]. Using STEEL, one can test simultaneously the accretion ratio and the pair fraction generated from a given stellar mass function and cosmology.

Any changes to the stellar mass estimates such as photometry, background subtraction, IMF, e.t.c. that affect the stellar mass function in a given cosmology will create a change in the SMHM relationship. Therefore, by the systematic propagation demonstrated in this work, any stellar mass estimation will create systematic differences in the pair fractions. With the techniques presented throughout this thesis, one could retrieve the systematic differences created in pair fraction under multiple ΛCDM cosmologies and for any given set of stellar mass functions. As a further test of the effectiveness of our methodology, we show showed that when inputting into STEEL the SMHM relation from the Illustris simulations, we are able to retrieve the pair fractions independently inferred from the same simulations.

In the era of wide and deep surveys, such as EUCLID, constraining a model using a single multi-epoch data set with consistent photometry will become a reality. The advantages of this are twofold: By tuning the SMHM relation to a given survey over a large range of redshifts, the growth of the stellar mass function over time can be tested against the implied satellite accretion and star formation rate as in Chapter 4. This can be seen as a consistency test of the cosmological model or the consistency of the stellar mass and/or star formation rate estimations. Secondly, as in Chapter 3, one can test if the

high redshift SMHM relation is consistent with the low redshift satellite distributions. Although the constraints on a given photometry, cosmological model, satellite evolution, star formation rate, e.t.c... are still not complete, STEEL will allow for non-physical results to be identified. Furthermore, STEEL can be used to compare current and future data sets that may use different stellar mass estimations.

5.5.2 Morphologies

Given the very promising results of STEEL in predicting satellite number densities in different environments and epochs, we here took a step further and explored whether STEEL's cumulative number of major mergers is able to account for the local fraction of elliptical galaxies, and also explored the viability of several lenticular formation models. A firm understanding of how mergers affect the galaxy population is desirable to allow the use of sophisticated measures observed galaxies such as the concentration, asymmetry, and clumpiness to become tracers of galaxy formation history [198].
Despite the noticeably good agreement between model predictions and data in Figure 5.11, we stress that different input SMHM relations can, as shown in Chapter 4, substantially affect the accretion rate which in turn will modify the number of galaxies experiencing major mergers. It follows that any cosmological galaxy evolution model that uses mergers as a physical driver for galaxy transformation should first simultaneously and self-consistently closely reproduce stellar mass functions, the SMHM relation, satellite distributions, and pair fractions at previous redshifts.

5.6 Conclusions

In the first part of this Chapter, we showed that the input SMHM relations, based on different stellar mass estimations, have a significant impact on the predicted galaxy pair fractions. In short, the steeper the relation, the lower the predicted pair fraction. Specifically, we compare stellar mass functions created and a de Vaucoulers-based photometry (cmodel) to a Sérsic-Exponential photometry (PyMorph), the latter leading to an enhancement in the number density of high mass galaxies. The effect of these stellar mass functions is a different input SMHM relation to STEEL, the primary difference consisting of a steeper high-mass slope when adopting the Sérsic-Exponential profile. As expected, the Sérsic-Exponential results in a lower pair-fraction. To attempt to explain the difference in pair-fraction evolution with redshift, we create a suite of toy models testing different alterations to the SMHM relation. We find that this evolution is linked to the evolution of the high-mass slope.
The purpose of this work is to show how subtle changes in the derivation of stellar mass could lead to large differences in the observationally-calibrated pair fractions. It is therefore crucial when comparing two different samples, to account for the systematic biases

introduced by the different assumptions implicit in stellar mass estimation. Observations or models that do not accurately match the stellar mass function measured in the observational samples they compare to, will not be able to self-consistently reproduce results.

The second part of this Chapter focuses on the implementation and validation of popular morphological models into STEEL. STEEL then calculates the models expected outcome on the statistically complete galaxy population. We found that the major merger model, according to which elliptical galaxies are foremost formed via mergers of nearly equal mass progenitors, well reproduces the observed local distribution in SDSS. Whereas, the Cook et al. [5] model for lenticular formation does not well reproduce the observed morphological fractions. This ability to flexibly incorporate models without the need to substantially alter the core of the simulation is what enables the transparency of semi-empirical models. Our minimal approach allows to gradually include in SEMs different galaxy physics and complexity, with the least possible assumptions and parameters, allowing for extreme transparency, control, and scrutiny of the outputs.

In conclusion, future surveys should look to use fast and flexible modelling such as STEEL, alongside data to be able to properly probe the systematic effects of assumptions that have been made to derive data products. For example, a hierarchical galaxy model must simultaneously reproduce: traditional abundance matching such as the SMHM relation, self-consistency between satellite accretion and central galaxy growth, and the normalisation and evolution of the galaxy pair fraction. This multi-product fitting will ensure relations such as the SMHM relation are not only better constrained but can also be co-constrained with other observables. In addition to this, STEEL provides a platform that combines fast development and running and testing of a variety of theoretical models of galaxy evolution. It is in developments such as these that we see how the design principles as laid out in Chapter 2 fundamental to STEEL such as speed, flexibility, and transparency, begin to pay dividends.

Chapter 6

Conclusion

"We gaze continually at the world and it grows dull in our perceptions. Yet seen from another's vantage point, as if new, it may still take the breath away." Alan Moore - Watchmen

This thesis describes STEEL, the STastical sEmi-Empirical modeL, a model designed to use the empirical technique in a new fashion. The design of STEEL allows for extremely fast, flexible and transparent testing of a variety of galaxy formation theories. The novelty introduced by the statistical dark matter accretion histories allows STEEL to make predictions that would be limited with other traditional (discrete/non-statistical) numerical or analytic tools of galaxy formation. The foremost problem concerning galactic modelling on a cosmological scale is represented by the large number of physical processes at play at different scales and the limited knowledge in how to describe them. Complexity and uncertainty are built into models from all contributing aspects, from the dark matter simulations [e.g. 112], to observations [e.g. 14, 84, 120]. Techniques have been developed to try to reduce the impact of these complexities, for example:

- To reduce the need to build stellar mass from first principles following the gas collapse. Abundance matching is used to create a mapping between galaxy stellar mass at the centre of a halo and the mass of the host halo itself at any redshift accessible by observations. In doing so it provides a robust framework to model the evolutionary tracks of galaxies following the growth of their host dark matter haloes.

- To remove tensions between conflicting observations such as the cosmic stellar mass density and cosmic star formation rate. Continuity modelling is used to predict the star formation rate using the evolution of the stellar mass function from high to low redshift.

The impact of publication bias, the preferential publishing of positive and novel results, is a known problem in academic disciplines [e.g. 199]. It is unsurprising then that the

number of papers that report new techniques or fits to data far outweighs those models which deconstruct the techniques to understand how our models work and document their limitations [e.g. 71, 111, 112]. Additionally, it can be expected that, for example, tuning of high parameter systems to create high fidelity matches to data, will generate a higher publication impact than equally important analysis tools. The design of STEEL has been influenced by this culture. We prioritise the understanding of systematics and internal self-consistency above that of fitting new features. Where new features are introduced they use modular and flexible modelling techniques that can be analysed independently. STEEL is a statistical model and not a physical model, in this way is the antithesis of high-resolution single galaxy or cluster simulations such as FIRE [54]. Galaxy modelling is currently based on a large spectrum of techniques from hydrodynamical simulations, through semi-analytic and semi-empirical models, to mock catalogues produced via HOD. In this regime, STEEL can be thought of as occupying a space between traditional semi-empirical and HOD modelling. We retain the ability to track galaxy populations in redshift but forgo the tracking of discrete objects from step to step. Despite the antithetical nature of STEEL to the hydrodynamical and non-statistical models, it is not adversarial, used in conjunction with other techniques it will prove to be a valuable tool for the galaxy modelling community.

6.1 Pros and Cons of STEEL as a Galaxy Model

STEEL has had the following major successes:

- Reproduction of the statistical distribution of satellite galaxies in dark matter halos over a broad range of redshifts and halo masses, a challenge for traditional ΛCDM models. (Chapter 3)
- Identification of the inconsistencies between certain stellar mass functions and dark matter accretion histories produced by ΛCDM cosmologies. (Chapter 4)
- Derivation of the star formation rate from a new halo centric approach, consistent with cutting-edge observations.(Chapter 4)
- Analysis of the systematic effects the estimation of stellar-masses has on the observed galaxy pair fraction. (Chapter 5)

Each of the points listed above directly stems from the statistical approach of STEEL to dark matter accretion histories. However, this technique loses individual galaxies that could be tracked through hydrodynamical, semi-analytic, and traditional semi-empirical models.

STEEL is an exceptional tool for systematic modelling, by applying predictions to statistically significant volumes we are able to draw transparent conclusions about the way

assumptions and models propagate into observed populations. Furthermore, Chapter 5 shows the potential of this systematic technique to track how changes to input propagate to output in a complex system, as shown with the pair fractions. Additionally, STEEL allows for fast, effective, and precise testing of a variety of recipes for galaxy evolution, such as the formation of bulges via mergers and disc instabilities.

6.2 Impact of STEEL

6.2.1 Modelling

Creating another cosmological model acting on a discrete dark matter background has little value to add to the large number of competitive models already existing in empirical, [e.g. 2, 6, 17, 33], analytic [e.g. 1, 31, 76, 200], and hydrodynamical [e.g. 27, 54, 147] regimes. The systematic statistical approach of STEEL can be used to complement these models in two ways:

- Pre-Processing:
  - STEEL can be used to test the validity of the combinations of input data. For example as in Chapter 4 we combine the dark matter assembly histories with two different stellar mass functions to test if the satellite galaxy accretion is consistent with the central galaxy growth. We find only one of the two tested SMF to produce an internally self-consistent satellite accretion - central mass growth. By validating the input data before running the simulation we ensure the output validity in advance of resource/time commitment.
  - STEEL can rapidly test the ability of different theoretical models to reproduce key observables such as the morphological mix of galaxies in the local Universe or the star formation rate distributions.
- Cosmological extrapolation:
  - STEEL can generate accurate predictions for a given galaxy model over scales covering many orders of magnitude, unfeasible for current hydrodynamic simulations. For example, STEEL can produce the distributions of massive satellites over a large scale range from groups to very massive clusters, to high redshifts in good agreement with data. Whilst the predictions for the group-scale are in good agreement with the Illustris TNG simulation, the cluster-scale is a new achievement of STEEL, not present in the Illustris TNG simulation.

6.2.1.1 Data

We have shown how STEEL can be used to combine data with a given cosmology to: predict stellar-mass - dark matter assembly consistency, satellite distributions, and star

formation rates. This has been used to add weight to the claims made about light profiles from Bernardi et al. [120] and the star formation rates from Leja et al. [14]. In each of the aforementioned works, problems are identified in current data analysis and more advanced data fitting routines are used. Specifically, Bernardi et al. [120] introduced their PyMorph photometry algorithm which provides more robust fitting to the light distributions of especially the more massive galaxies and more careful subtraction of the sky background. Leja et al. [14] put forward a state-of-the-art Bayesian SED fitting model, Prospector-alpha, that better accounts for the emission from old stars, thus lowering the amount of younger stellar population and overall SFRs. The previous examples show that STEEL has a valuable contribution to make to the validation of data fits, fully realised it could become an integral part of data fitting routines bringing theory closer to observations.

6.3 Future of STEEL in Galactic Astrophysics

The future of STEEL and potentially the future of statistical semi-empirical modelling now relies on communication of the advantages (and disadvantages) of the technique. We have begun to build collaboration around the model with several other empirical modellers, dark matter physicists, observers (both survey designers and data analysts), as well modellers that use semi-analytic and hydrodynamic models. The reception to STEEL has been widely positive with many intrigued by its unique capabilities. The most notable group involved is the EUCLID consortium. Working with EUCLID, STEEL has the potential to provide firm constraints on the expected number of pair fractions, mergers, and ellipticals at different redshifts (one of the main aims of Work Package 5, for example). To provide the best performance, STEEL will need to input the same stellar mass reference system as adopted in the survey.

PJG outlined the methodology to create STEEL alongside a 4-year development plan for a fellowship proposal[1]. In this plan the future of STEEL was separated into two complementary pathways, a data fitting tool and a galaxy modelling tool, presented below:

6.3.0.1 Data fitting tool

HOD modelling is a popular analytic tool used to create galaxy mocks in terms of the average number of central and satellite galaxies above a certain luminosity/stellar mass threshold as a function of host halo mass. Using relationships such as HOD number counts, or a simple SMHM relation from abundance matching, can produce galaxy mocks onto a dark matter simulated light cone, which in turn can be used for predictions on

[1] Unfortunately unsuccessful, the proposal and related documents are included in full in Appendix E.

observed number counts above the flux limit of a survey. This kind of modelling is done prior to first light and far separated from the actual photons to be received. With STEEL we can outperform traditional mock modelling. STEEL can not only predict galaxy number counts for any given SMHM relation derived from measured SMFs, but it can also check for self-consistency in the measured SMFs and thus guide the overall initial fitting procedure. More specifically, STEEL can check if the total galaxy growths implied by the assumed SMF/SMHM relations are larger than the total stellar mass assembled via satellite accretion, a condition that, as we discussed in Chapter 4, can be violated in a ΛCDM Universe.

Using Figure 6.1 we show how STEEL can be positioned to become a key element of observational fitting bringing theory and observation much closer than ever before. The observational inputs (green) are twofold, the survey, i.e., the flux received on the telescope which is the only constant in the system, and the observational fits, which are used to derive physical attributes from the collected photons. In this example the survey and fits generate the stellar mass function (SMF). Secondly, the dark matter cosmology is input in STEEL[2]. As described throughout this thesis, STEEL uses halo mass functions (HMF), halo substructures, and halo growth histories which are then combined to generate a statistical dark matter accretion history. Most of the baryonic (SMF) and DM (accretion histories) inputs described above are flexible and fast to model. However, the final modelling usually adopted to interpret the data is traditionally computationally intensive, time-consuming, and often degenerate. The empirical approach, especially as used within STEEL, offers solutions to these problems being computationally light, fast, and transparent. As described in Chapter 4, using Figure 4.4 we can derived the star formation rate and bimodality using a combination of observation, cosmology, and modelling. Through comparison of modelled properties to observations, we can test that all elements of theory and data used are self-consistent and evolve appropriately together within a given cosmology e.g.

- Satellite galaxy distribution
- Pair fraction
- Star formation rate / cosmological star formation density
- Star formation - stellar mass function continuity
- Galaxy bimodality
- Galaxy morphology

Items closer to the top of the list are less affected by the modelling and therefore likely to be more robust within this method. Given empirical techniques use observation as inputs

[2]it is worthy of note that the exact cosmological parameters may also influence the fits used.

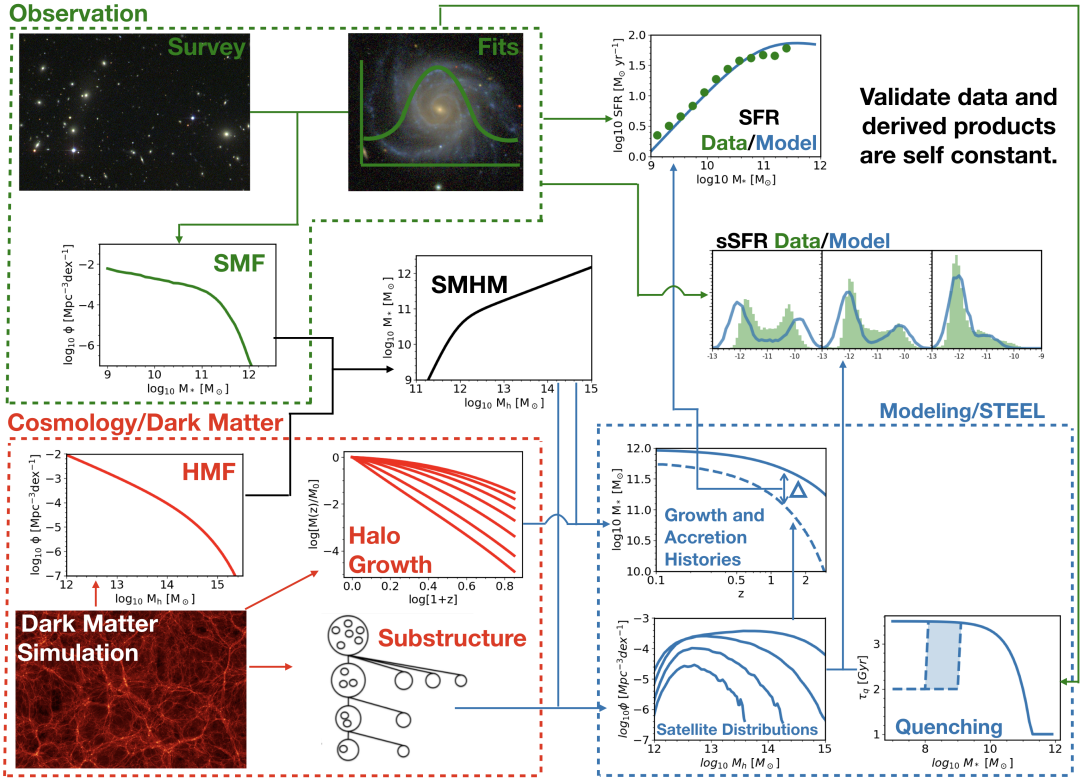


FIGURE 6.1: Schematic cartoon related to Figure 4.4 of how STEEL can be used to check for self-consistency between fitting models used on observed data and the theoretical assumptions that underpin the fitting models.

to modelling, having an empirical model as part of a data fitting pipeline the model will have the potential to reject/modify its own inputs by influencing the parametrisation of the observational data reduction routines. This fitting, checking, refitting, cycles will ensure that any mismatch between the observational fit and the modelled property can be addressed from multiple distinct angles appreciating the uncertainties inherent in fitting, modelling, and cosmology together.

6.3.1 Galaxy modelling tool

STEEL, built around the concept of a statistical dark matter accretion history, can be regarded as a new competitive tool in the modelling of galaxies. This is due to two factors: firstly, the ability to simulate without limits on volume or resolution, thus allowing to model even the rarest objects in the Universe, a feature that is not realistically achievable by other modelling techniques; secondly, the removal of discrete objects allows for a new type of continuity check where the consistency of populations is modelled as a priority instead of the consistency of individual haloes/galaxies.

STEEL is an excellent start to statistical semi-empirical modelling. With any future revision, the model should be updated in the following ways:

- The statistical dark matter accretion history should be a separate library, containing flexible cosmology, halo definitions (e.g. virial, splash-back, m200c, e.t.c...), and should be built in an object-oriented fashion such that it is easily extendable.
- The galaxy modelling should be redesigned such that each dark matter accretion history is built upon using explicit probability distribution convolutions.
- The outputs should be shaped by the statistical dark matter accretion history and contain structured probability distributions to be able to post-process a hierarchical complexity of results.
- The post-processing suit should contain any and all integrators required to process the outputs into galaxy populations for cosmological volumes, single mass tracks, or given mass ranges.
- The post-processing suit should offer 'views' on the data structure using flexible plotting routines for ease of query and graph creation by a non-expert user similar to the API provided to IllustrisTNG [23, https://www.tng-project.org/data/vis/] as well as a more complex return structure to extract the full data products.

As detailed in the PERT chart in Appendix E, this would take six months to one year of full-time development depending on the programming and empirical modelling skills of the developer. Once developed, documented, and deployed such a tool would be of enormous value to the galactic astrophysics community.

6.4 In closing

The applications of STEEL detailed in Section 6.3 each independently represent a tool that, if correctly developed and made publicly available, would be at the cutting edge of galactic astrophysics. The work done in this thesis represents the foundations upon which these tools can be built.
In brief, the findings from this thesis can be summarised as follows:

- The dark matter structure and dynamics is the first-order driver for prediction of galaxy properties including, satellite distributions, star formation rate, galaxy morphologies.
- Assumptions made during stellar-mass estimation produce significant non-trivial systematics in the pair fraction of galaxies and likely further properties, and without proper appreciation of such systematics, we are likely to fundamentally misinterpret the nature of galaxy assembly.

- Stellar mass functions with an enhanced number densities of high mass galaxies are in much better agreement with ΛCDM cosmology than those with lower number densities of high mass galaxies. Furthermore, some traditional stellar mass functions are fundamentally incompatible with ΛCDM hierarchical cosmology.

To our knowledge, the complexities of the systematics behind galaxy pair fractions has never been addressed before in terms of the input SMHM relation. Finally, the lack of self-consistency between hierarchical ΛCDM assembly and some stellar mass functions is a result of fundamental importance to galactic astrophysics that it was possible to achieve thanks to the statistical power of STEEL. This result in particular hints at a potentially devastating departure between observation and theory that, if left unaddressed, will hinder our understanding of the evolution of galaxies for the foreseeable future.

Appendix A

Additional Semi-Analytic Methods

In this Appendix, we detail a further common consideration taken in semi-analytic modelling, the state of the gas. The state can be calculated in several different ways, an empirical pressure based relationship is given by Blitz and Rosolowsky [201], the pressure in the disk is related to the ratio of molecular and atomic hydrogen.

$$R_{\mathrm{H}_2} = \left(\frac{\Sigma_{\mathrm{H}_2}}{\Sigma_{\mathrm{HI}}}\right) = \left(\frac{P_m}{P_0}\right)^{\alpha} \tag{A.1}$$

The H_1 & H_2 surface densities are given by Σ_{HI} & Σ_{H_2}, P_m is the mid disk pressure, and P_0 & α are additional free parameters. The gas partitioning can be calculated though an analytic model based on the connection between the interstellar radiation filed and the molecular self shielding [202–204],

$$f_{H_2} = 1 - \left[1 + \left(\frac{3}{4}\frac{s}{1+\delta}\right)^{-5}\right]^{-1/5} \tag{A.2}$$

.

This is by no means a complete record of the various analytic recipes used in semi-analytic modelling. However here we have shown a subset of the multitude of free parameters that enable tuning of such models to achieve results consistent with observations.

$$s = \ln(1 + 0.6\chi)/\left(0.04\Sigma_{\mathrm{comp},0}Z'\right), \tag{A.3}$$

$$\delta = 0.0712\left(0.1s^{-1} + 0.675\right)^{-2.8}, \tag{A.4}$$

$$\chi = 0.77\left(1 + 3.1Z'^{0.365}\right), \tag{A.5}$$

where Σ_{comp} is the surface density for a given 100pc atomic-molecular cloud.
The final mechanism we discuss here is the enrichment of the galaxy and halo gas with metals. During the process of star formation and stellar mass, recycling metals are created. This is modelled as a batch process where $\mathrm{d}M_Z = y\mathrm{d}m_*$ where a mass $\mathrm{d}M_Z$ of

metals is produced in each batch of star formation dm_*, y is a free parameter. The metal-enriched gas formed in this process is ejected by supernovae and assume instantaneously mixed with the cold disk gas. As supernovae are thought to be one of the main drivers of galactic wind the metals are thought to be preferentially ejected with the wind parameter ζ controls the ejected metal fraction, and the equation for the metal mass in the galaxy is updated as such,

$$\zeta = \zeta_{\rm lo} \exp\left(-M_h/M_{\rm ret}\right), \tag{A.6}$$

$$\dot{M}_Z = y(1-R)(1-\zeta)\dot{m}_* + Z_{\rm hot}\,\dot{m}_{\rm inf} - Z_{\rm cold}\,\dot{m}_{\rm out}\,, \tag{A.7}$$

ζ_{10} and $M_{\rm ret}$ are free parameters, R is the recycled fraction and $Z_{\rm cold}$ is the metallicity of the cold gas.

Appendix B

Stellar Mass Assembly: Comparison to other models.

In Figure B.1 we show the in-situ vs ex-situ growth with the same model as shown in Figure 4.3, we add to this plot data extracted from the Illustris TNG100 simulation. In Figure 5.4 we see Illustris has a shallower slow mass slope and a steeper high mass slope such that more stellar mass is mapped into haloes of all sizes. We see the change in both of these slopes reflected in the accretion histories, firstly, for the lower mass galaxies (see $log10M_{*,cen} = 10^{11}$) closer to the SMHM knee we find enhanced accretion due to the larger masses from more minor mergers. Secondly, the high mass slope is a direct result of the accretion, to support the same merger assembly with the higher mass galaxies in the satellite haloes above the knee where galaxy growth is dominated by the accretion the galaxy growth with halo size must be enhanced.

In Figure B.2 we show the in-situ vs ex-situ growth with the same model as shown in Figure 4.3, we add to this plot data from the EMERGE model from Moster et al. [17] shown as black lines. The solid lines show the total galaxy mass followed back selecting populations by mass at redshift $z = 0.1$. The dotted and dashed lines show the amount of galaxy mass formed in-situ and ex-situ respectively. In all cases EMERGE predicts satellite accretion becomes the dominant mass growth pathway at higher redshifts then STEEL. In the third column we see that EMERGE and STEEL also disagree about the mass growth history of $log10M_{*,cen} = 11$ galaxies, however, both models agree that the dominant mass growth path of galaxies at this mass are in-situ processes.

In Figure B.3 we show for the $log10M_{*,cen} = 11.5$, and 11 galaxies the central growth and star formation rate ratio from Behroozi et al. [6]. The central growth is close to that found from STEEL and the star formation rate transition for $log10M_{*,cen} = 11.5$ is an excellent match.

In Figure B.4 we show a comparison with the Semi-Analytic model described in Menci et al. [18]. At all masses the stellar growth is substantially different from STEEL and

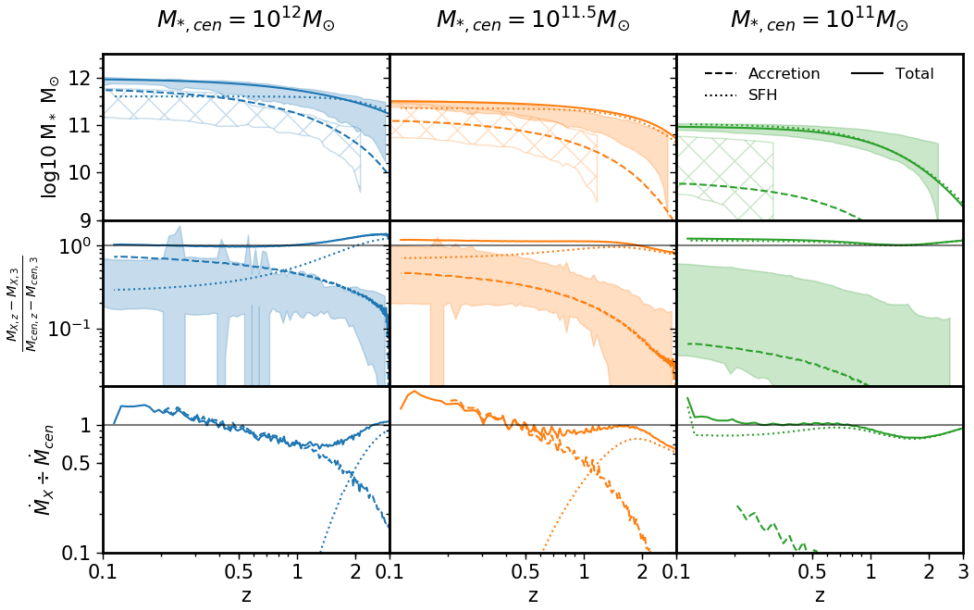


FIGURE B.1: As in Figure 4.3 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ in blue orange and green respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies the star formation rate and the mass growth rate of the central galaxy predicted by abundance matching. In the top panel, the shaded regions are galaxies selected from the Illustris simulation the hashed region is then the satellite accretion from Illustris, in the middle panel the shaded region is the ratio of satellite accretion from Illustris. The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should, therefore, be close to or on these lines.

the other models shown in this appendix. Furthermore, the Semi-Analytic model shows little change in the accreted mass ratio over cosmic time, again this is inconsistent with the findings from STEEL and the other models presented in this section. We attribute most of the differences seen here to the substantial difference in the SMHM relationship between the two models.

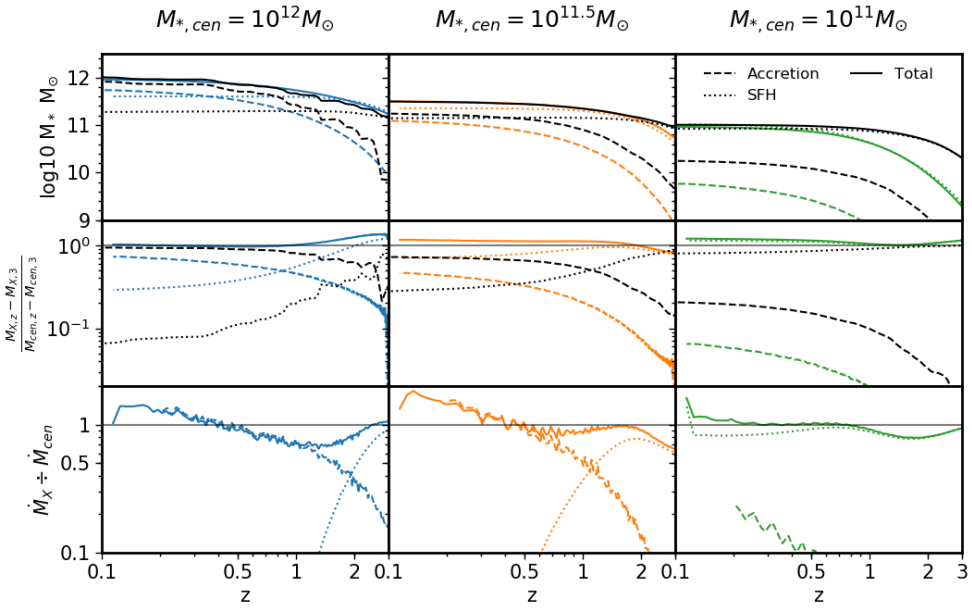


FIGURE B.2: As in Figure 4.3 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ in blue orange and green respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies the star formation rate and the mass growth rate of the central galaxy predicted by abundance matching. In the top and middle rows we add black lines to show the in-situ and ex-situ growth from EMERGE Moster et al. [17]. The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should, therefore, be close to or on these lines.

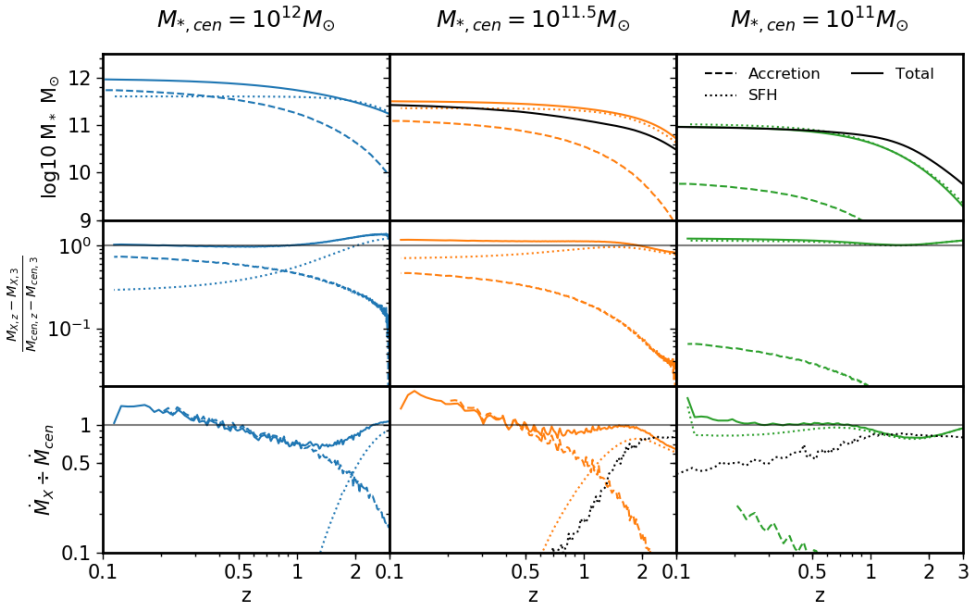


FIGURE B.3: As in Figure 4.3 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ in blue orange and green respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies the star formation rate and the mass growth rate of the central galaxy predicted by abundance matching. In the top and bottom rows, for the $log10M_{*,cen} = 11.5$, and 11, we add black lines to show the central galaxy growth and the star formation rate ratio from Behroozi et al. [6]. The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should, therefore, be close to or on these lines.

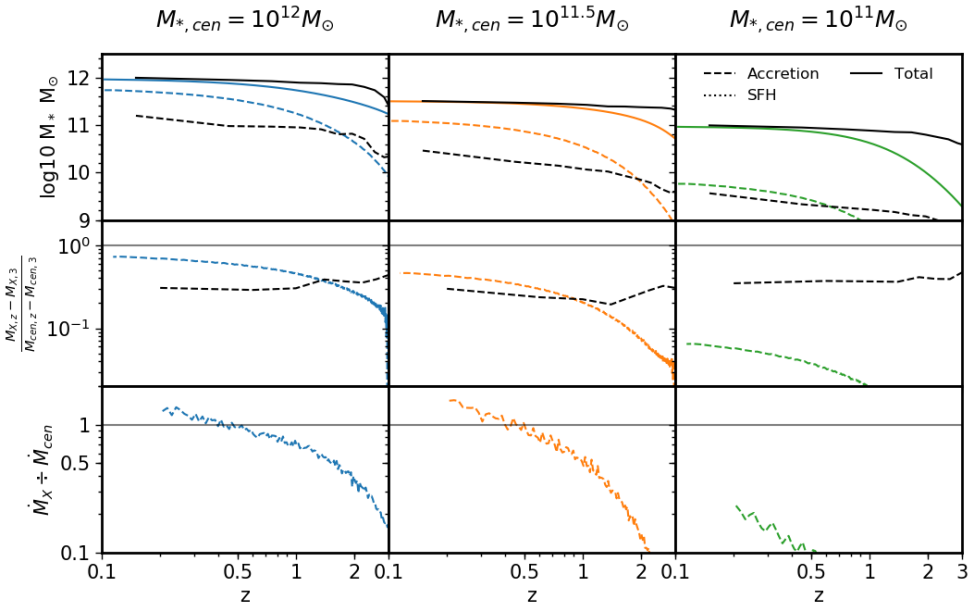


FIGURE B.4: As in Figure 4.3 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ in blue orange and green respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies the star formation rate and the mass growth rate of the central galaxy predicted by abundance matching. In the top and middle rows, we add black lines to show the ex-situ growth and central growth from the Semi-Analytic model described in Menci et al. [18]. The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should, therefore, be close to or on these lines.

Appendix C

Parameter cross-sections from abundance matching MCMC.

Figure C.1 shows the redshift $z = 0.1$ output from the MCMC abundance matching fits. It becomes immediately obvious that the low mass slope (β) is poorly constrained however the impact on the SMF is limited within the margin of error. The position of the knee (M) is well constrained against both the normalisation (N) and the high mass slope (γ). The shape of the constraint between the normalisation and gamma emanates from the need to produce high mass galaxies, if the normalisation is decreased the slope must increase to ensure enough haloes produce massive galaxies.
Figure C.2 shows the redshift $z > 0.1$ output from the MCMC abundance matching fits. All parameters have low evolution and the SMHM relation evolves only weakly with redshift. For M, β, and γ where the distributions are wide or close to the prior we have tested wider priors and insignificant change is found.

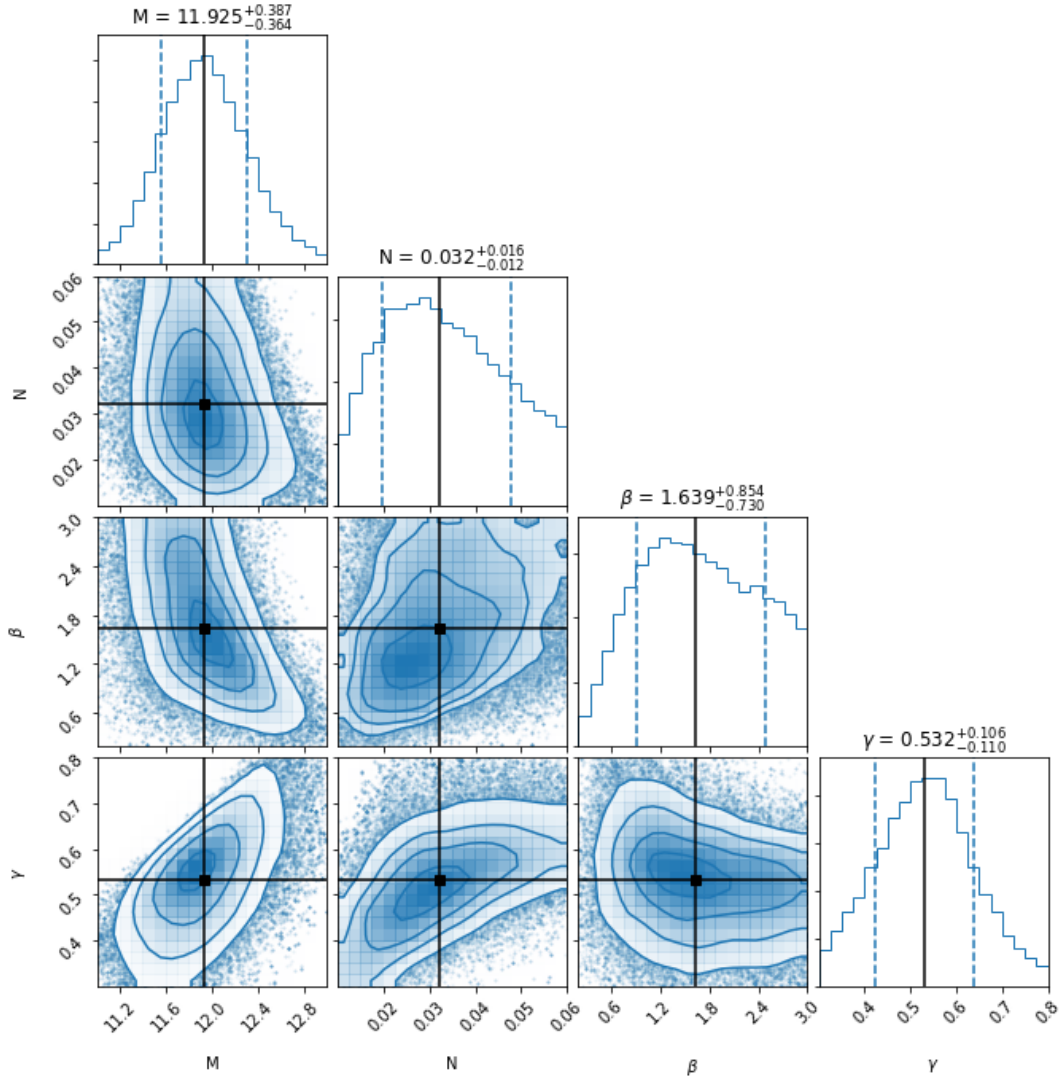


FIGURE C.1: We show the MCMC parameter space for the redshift $z = 0.1$ fit. The position of the knee (M), the normalisation (N) the low mass slope (β) and the high mass slope (γ) are shown from left to right. Columns are titled with the best fit values and 16th/84th percentile errors. The black lines show the best fit value with a black square at intersections, the 16th/84th percentiles are shown with blue dashed lines on the histograms.

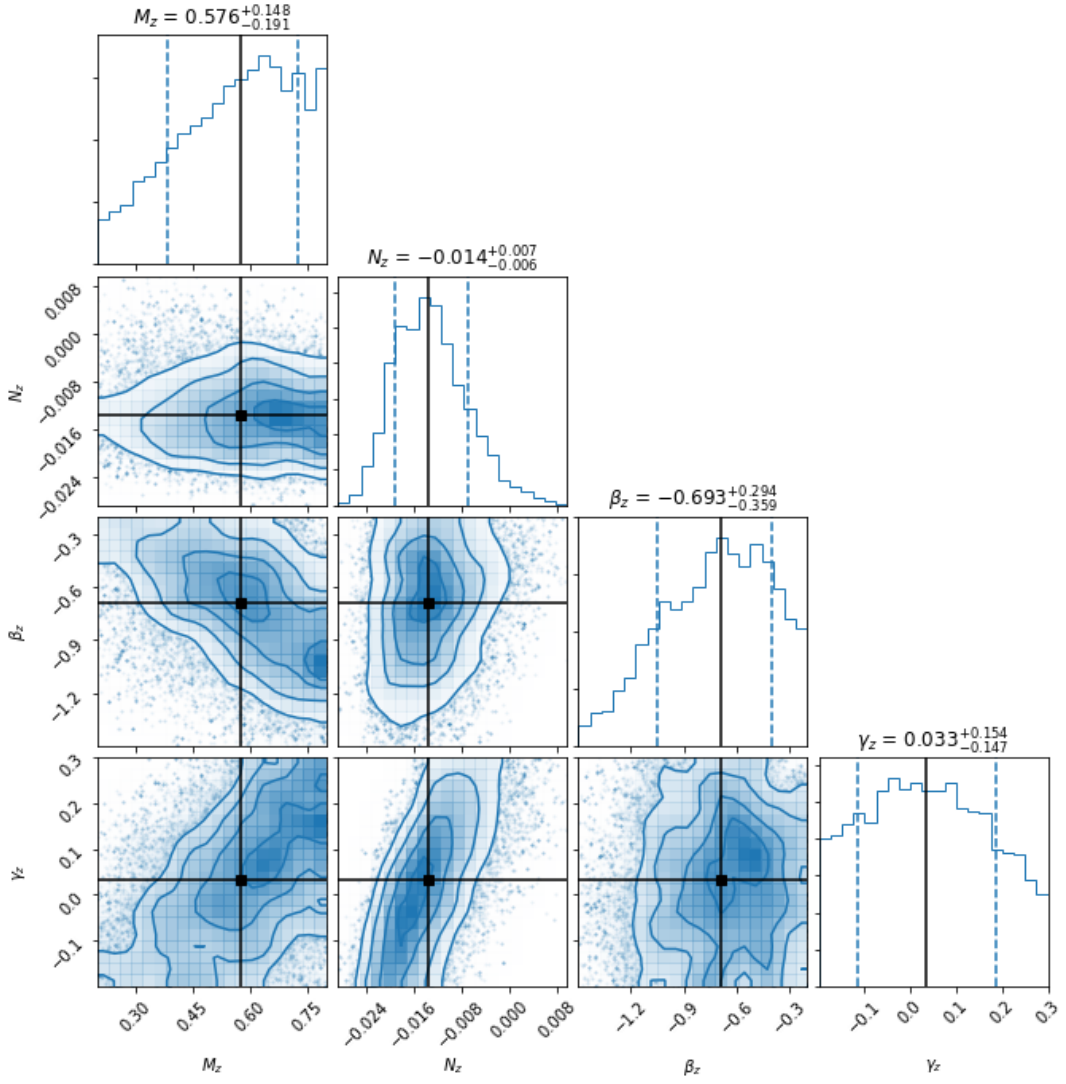


FIGURE C.2: We show the MCMC parameter space for the high redshift $z > 0.1$ fit. The evolution of: the position of the knee (M_z), the normalisation (N_z) the low mass slope (β_z) and the high mass slope (γ_z) are shown from left to right. Columns are titled with the best fit values and 16th/84th percentile errors. The black lines show the best fit value with a black square at intersections, the 16th/84th percentiles are shown with blue dashed lines on the histograms.

Appendix D

Full publication texts

The PDFs of the three published papers are provided in full in order of publication. Papers one [19] and two [20] are the prints from MNRAS Paper 3 [21] is the preprint from arXiv.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **491,** 634–654 (2020) doi:10.1093/mnras/stz2956
Advance Access publication 2019 October 22

Predicting fully self-consistent satellite richness, galaxy growth, and star formation rates from the STatistical sEmi-Empirical modeL STEEL

Philip J. Grylls [1]★ F. Shankar,[1] J. Leja [2] N. Menci,[3] B. Moster [4] P. Behroozi [5] and L. Zanisi[1]

[1]*Department for Physics and Astronomy, University of Southampton, Highfield SO171BJ, UK*
[2]*Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, MA 02138, USA*
[3]*INAF – Osservatorio Astronomico di Roma, via di Frascati 33, I-00040 Monte Porzio Catone, Italy*
[4]*Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr 1, D-81679 Müchen, Germany*
[5]*Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ 85721, USA*

Accepted 2019 October 18. Received 2019 October 14; in original form 2019 May 2

ABSTRACT

Observational systematics complicate comparisons with theoretical models limiting understanding of galaxy evolution. In particular, different empirical determinations of the stellar mass function imply distinct mappings between the galaxy and halo masses, leading to diverse galaxy evolutionary tracks. Using our state-of-the-art STatistical sEmi-Empirical modeL, STEEL, we show fully self-consistent models capable of generating galaxy growth histories that simultaneously and closely agree with the latest data on satellite richness and star formation rates at multiple redshifts and environments. Central galaxy histories are generated using the central halo mass tracks from state-of-the-art statistical dark matter accretion histories coupled to abundance matching routines. We show that too flat high-mass slopes in the input stellar mass–halo mass relations as predicted by previous works, imply non-physical stellar mass growth histories weaker than those implied by satellite accretion alone. Our best-fitting models reproduce the satellite distributions at the largest masses and highest redshifts probed, the latest data on star formation rates and its bimodality in the local Universe, and the correct fraction of ellipticals. Our results are important to predict robust and self-consistent stellar mass–halo mass relations and to generate reliable galaxy mock catalogues for the next generations of extragalactic surveys such as Euclid and LSST.

Key words: Galaxy: halo – galaxies: abundances – galaxies: clusters: general – galaxies: evolution – galaxies: high-redshift – galaxies: star formation.

1 INTRODUCTION

Galaxies are thought to grow and evolve through a combination of '*in situ*' and '*ex situ*' processes. *In situ* processes such as star formation, are thought to be driven by the availability of cold gas in the galaxy. The reserve of cold gas ready to fuel star formation could be regulated by a number of internal and external processes, from stellar and active galactic nuclei feedback to host halo and/or morphological quenching (e.g. Granato et al. 2004; Dekel et al. 2009; Lilly et al. 2013; Schawinski et al. 2014). One important *ex situ* channel affecting galaxy growth is satellite accretion. In particular, in very massive galaxies growth via satellite accretion has been claimed to become progressively more relevant (De Lucia et al. 2006; van Dokkum et al. 2010; Shankar et al. 2013, 2015; Buchan & Shankar 2016; Groenewald et al. 2017; Matharu et al. 2019). Central galaxies that reside at the centre of massive haloes thus provide a window into the different pathways that have contributed to the mass growth history of galaxies in the local universe. Exploring the way these galaxies build their mass can give insights into the stellar mass–halo mass (SMHM hereafter) relation, the efficiency of the satellite transport from the edge of the cluster to the centre, the balance of the major processes taking place on these satellites, the galaxy merger rate, and the star formation rate (SFR). The characteristic mass at which galaxies transition from being *in situ* to *ex situ* growth dominated has previously been found at $M_* \sim 10^{11}$ M$_\odot$ (Bernardi et al. 2011; Cattaneo et al. 2011; Shankar et al. 2013).

Models of galaxy formation traditionally use the hierarchical growth of dark matter structure as the backbone for galaxy assembly. Hydrodynamical simulations co-evolve the dark matter and baryonic matter allowing for a simultaneous look at the assembly of both components (Vogelsberger et al. 2014; McAlpine et al. 2015). The latter technique, however, requires large computational

★ E-mail: p.grylls@soton.ac.uk

resources. Less computationally intensive models such as traditional semi-analytic and semi-empirical models use dark matter merger trees from post-processing of dark matter simulations (Guo et al. 2011; Shankar 2013). Dark matter merger trees visualize dark matter assembly as a central trunk and halo mergers happen where branches join. Semi-analytic models initialize gas at high redshift and use a number of physical assumptions and free parameters to tune to observations (De Lucia et al. 2006; Guo et al. 2011). Semi-empirical models use a more direct approach initializing galaxy stellar mass in dark matter haloes most commonly through abundance matching, the association of galaxies to dark matter host haloes via relative abundances (Hopkins et al. 2010b; Zavala et al. 2012; Moster, Naab & White 2013; Shankar et al. 2014; Moster, Naab & White 2018). Both semi-empirical and semi-analytic models follow the merging histories of the underling dark matter merger trees to track the *in situ* and *ex situ* build-up of galaxy mass. The work of Moster et al. (2018), for example, uses a semi-empirical model to associate the growth of the dark matter halo to the SFR of the host galaxy alongside the build-up of stellar mass from satellites accretion, further strengthening the connection between the dark matter host environment and the build-up of galactic stellar mass.

An average measure of the growth of galaxies can be obtained by comparing stellar mass functions (SMFs), the number densities of galaxies as a function of mass, over multiple epochs. Selecting galaxy populations with the same number density at each epoch, assuming that galaxies maintain rank ordering over cosmic time, allows an estimation of the average growth of galaxies to be made. An estimation of the SFR for each galaxy population can then be computed by taking the time derivative of the stellar mass growth. However, the star formation generated in this way is significantly lower than observational estimates of the SFR (e.g. Leja et al. 2015; Lapi et al. 2017). It is consequently found that if observed SFRs are used in models, they cannot be reconciled with the SMF. In Grylls et al. (2019) it is shown this is also in effect in satellite galaxy distributions, where the predicted number of massive satellites is far too high if satellites evolve using observed SFRs. This is a particular problem for semi-empirical models where one would ideally use the observed SFR as an input. To overcome the inconsistencies between observed SFRs and model predictions it is possible to include SFRs generated by the method above commonly referred to as a continuity SFR. Attributing the stellar mass growth to star formation in this way yields an upper limit to SFR that is consistent with the SMF evolution by design.

To properly constrain the formation of a galaxy one must reproduce the galaxy environment, the distribution of satellites around the central galaxy, at all previous redshifts. Discrepancies with observations of the high-redshift environment will cause modelled satellite stellar mass accretion rates that are either too high or too low. To account for such deficit/surplus modelled *in situ* growth must compensate through other modelling parameters to maintain the evolution of the stellar mass density. Such compensation could, for example, be of the form of suppressed/enhanced SFR or alternatively any number of other physical modelling parameters. Reproducing the number density and distribution of galaxies has, however, proven a challenge for many semi-analytic models (e.g. Asquith et al. 2018). Furthermore, where semi-analytic models have included more physics via an increased number of modelling parameters, this has led to degeneracies that obscure which are the essential physical processes governing galaxy formation (e.g. González et al. 2011; Lapi et al. 2011). Semi-empirical models, due to the direct initialization of galaxies to haloes and smaller parameter spaces, fare better as they can by design provide more clarity as to which modelling assumptions and related parameters are necessary to fit observations.

In our previous work we presented STEEL (Grylls et al. 2019, hereafter referred to as Paper I), a STatistical sEmi-Empirical modeL. The basis of STEEL was the shift from discrete merger trees in favour of statistical halo growths and merging histories, which enables to probe galaxy environment unbiased by volume and mass resolution. In Paper I we presented a detailed study of the richness of the galaxy group and cluster environments in the local Universe. In this work we extend the analysis of satellite richness from STEEL to high redshifts comparing with a large galaxy cluster survey, and state-of-the-art hydrodynamical simulations. Having a clear and well-constrained picture of the building up of satellite population then allows STEEL to create more reliable merger histories for central galaxies across cosmic time. Using STEEL's improved merger histories we are then able to constrain the *ex situ* growth and, by extension, derive more reliable estimates of the *in situ* growth and the implied SFRs of central galaxies.

The paper is laid out as follows: In Section 2 we present the halo and stellar mass functions we use for our abundance matching as well as the high-redshift clusters we adopt to constrain the model performance. In Section 3 we discuss STEEL, most importantly we provide an overview of the statistical merging history in Section 3.1, the abundance matching in Section 3.3, as well as updates made to the model described in Paper I. We begin the results by testing the high-redshift halo merger rate in Section 4.1, in Section 4.2 we then present the high-redshift satellite galaxy distribution results, and in Section 4.3 the growth of our galaxy population via *in situ* and *ex situ* processes. We then discuss our results in a wider context in Section 5 and summarize in Section 6.

2 DATA

In this section we first describe the simulations used for the halo mass functions (HMFs), then the data used to create the SMFs. Together, the HMF and SMF are used to create an SMHM relation, described in Section 3.3, which defines the galaxy–halo connection, essential to STEEL. We then provide details of the cluster data we compare to at high redshifts. All the data presented in this section are converted, wherever necessary, to a Chabrier (2003) stellar initial mass function (IMF). In this work we adopt the Planck cosmology with $(\Omega_m, \Omega_\Lambda, \Omega_b, h, n, \sigma_8) = (0.31, 0.69, 0.05, 0.68, 0.97, 0.82)$[1] (Planck Collaboration 2015). Halo masses are defined as virial masses in this cosmology, unless stated otherwise.

2.1 Halo mass functions

In this work we use the HMF from the simulations of Despali et al. (2016), generated and converted to appropriate units and cosmology using COLOSSUS (Diemer 2017). The HMF provides the number density of haloes in a given mass bin at a given redshift. We generate the substructure of subhaloes using the subhalo mass functions found in Jiang & van den Bosch (2016). The subhalo mass function provides the number density of subhaloes expected for a parent halo

[1]We note that Planck's best-fitting cosmology has slightly different parameters than those adopted in some of the observations used in this work, such as the SMFs ($(\Omega_m, h = (0.30, 0.70)$). Correcting the SMF volumes and luminosities to the same cosmology yields essentially indistinguishable results.

of a given mass. The (sub) halo mass functions used in this work are all calibrated against the Bolshoi Simulation (Klypin et al. 2016).

2.2 Stellar mass functions

In this work we use the SMFs defined below, along with the HMF given above, to constrain the SMHM relationship. The latter in STEEL is constrained first at low redshift $z = 0.1$, using SMFs from the Sloan Digital Sky Survey (2.2.1). The evolution of the SMHM relation is then constrained to match the SMF at higher redshifts ($z > 0.1$).

2.2.1 Low redshift, $z = 0.1$

At low redshift we use the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7) from Meert, Vikram & Bernardi (2015). The data from the SDSS-DR7 spectroscopic sample (Abazajian et al. 2009) contain $\sim 670\,000$ galaxies fitted with a Sérsic + exponential model (PYMORPH; Meert et al. 2015) with associated halo masses and central satellite classifications from Yang et al. (2012). The improved photometric analysis by Meert et al. (2015) provides more reliable estimates of the SMF at the high-mass end, which appear more abundant than previous estimates (Bernardi et al. 2016, 2017). In this work we investigate the effect of this enhanced high-mass end on galaxy assembly. We compare to previous determinations of the SMF using as an example the de Vaucoulers (de Vaucouleurs 1948) based cmodel fits from SDSS (Abazajian et al. 2009). The latter definition of galaxy stellar mass has been extensively discussed not to be accurate, partially due to incorrect sky subtraction and adoption of non-ideal light profiles (Bernardi et al. 2013). Bernardi et al. (2017) have clearly shown that the choice of light profile is not a simple matter of 'semantics'. The single or double Sèrsic models perform better in fitting the surface brightness of galaxies independently of the galactic environment (Meert et al. 2015). The performance is thus not related to the inclusion of the intragroup or intracluster light in the fit (Bernardi et al. 2017).

2.2.2 High redshift, $z > 0.1$

At higher redshift ($0.3 < z < 3.3$) we use SMFs from the COSMOS2015 catalogue (Davidzon et al. 2017). Here masses are defined using spectral energy distribution fitting, including ultradeep infrared photometry. Davidzon et al. (2017) use Bruzual & Charlot (2003) stellar population synthesis models to estimate stellar masses. As SED fitting is notably different from light profile fitting, one cannot apply the same corrections as in Mendel et al. (2014). Nevertheless, to match the mass-to-light ratios adopted by Mendel et al. (2014), based on the Bell et al. (2003) mass-to-light ratios, we follow Bernardi et al. (2013) and increase the Davidzon et al. (2017) stellar masses, based on Bruzual & Charlot (2003), by $+0.15$ dex. We note that the resulting $z = 0.37$ SMF after this correction is in remarkable good agreement with the $z = 0.1$ SMF by Bernardi et al. (2013). Our result also matches the findings by Bernardi et al. (2016), who showed that, by making use of the BOSS sample, the SMF shows negligible number density evolution up to $z \sim 0.5$.

2.3 Clusters

2.3.1 Cluster at $z = 2.5$, Wang et al. 2016

The highest redshift cluster we compare to is an $M_{\rm vir} = 10^{13.7}\,{\rm M}_\odot$ halo containing 15 galaxies with $M_* > 10^{10}\,{\rm M}_\odot$ at a redshift of $z = 2.5$. This cluster is reported in Wang et al. (2016), and we provide a brief description of the observation and data here. The cluster is observed using IRAM-NOEMA, VLT-KMOS, VLA, *XMM–Newton*, and *Chandra* for the spectroscopic observation and redshift determination. The galaxy masses are determined assuming a Salpeter (1955) IMF, which we correct to a Chabrier (2003) IMF, by decreasing the stellar masses by 0.24 dex. The halo mass ($M_{\rm vir} \sim 10^{13.93}\,{\rm M}_\odot$) of the cluster is estimated in three different ways, using the total X-ray luminosity, the velocity dispersion of its member galaxies above $M_* = 10^{10.76}\,{\rm M}_\odot$, and the stellar richness of the cluster.[2] Given this object was a targeted cluster, we cannot estimate the cosmic abundance (i.e. the number per cubic megaparsec). For analysis and comparison later in this work we assign this cluster an abundance of $N(>M_* = 10^{13.93}) = 10^{-7.15}$ (Mpc^{-3}) which is estimated by integrating the HMF in the limits $[10^{13.93}, \infty]$, thus providing an upper limit to the number densities associated with clusters of this mass.

2.3.2 1959 clusters at $z = 0.7$–1.0, Wen & Han 2018

We compare to the cluster sample from Wen & Han (2018), which contains 1959 clusters from SDSS-DR14 (Abolfathi et al. 2017) and the WISE survey (Wright et al. 2010). The clusters are identified in the *W1* band, and foreground objects are removed using the SDSS photometric data. The cluster mass and richness are estimated using the total *W1* band luminosity within 1 Mpc of the central galaxy. As performed above, to each cluster we assign an upper limit to their abundances from the cumulative integration of the HMF.

3 METHOD: STEEL

Our model STEEL is a STatistical sEmi-Empirical modeL designed to investigate the satellites and subhaloes in groups and clusters. In brief, STEEL removes reliance on discrete dark matter simulations or halo merger trees, commonly used in galaxy simulations, in favour of using mass functions to create a 'statistical dark matter accretion history' described in more detail in Section 3.1. This statistical history is then combined with semi-empirical techniques, such as abundance matching (Section 3.3), to create average galaxy population statistics. Whilst a comprehensive description of STEEL can be found in Paper I, we provide in this section highlights of STEEL, including any relevant updates. In this paper we have two objectives. First, we investigate the impact of different SMHM relations on the accretion histories of galaxies. The second aim is to use our semi-empirical accretion histories and galaxy growth histories to derive the SFR/star formation histories of galaxies.

The image in Fig. 1 shows a simple visualization of the process we use to determine the effect of photometry on the accretion histories of galaxy populations. Starting on the left we show two SMFs, the primary difference is the blue (dotted) SMF has a substantially enhanced high-mass end. In the middle panel we show how this high-mass slope changes the SMHM relation, an enhanced high-mass end SMF results in an enhanced high-mass slope. The galaxy growth histories, shown as solid lines, are generated using the SMHM relation used to calculate the average satellite stellar

[2]We note the velocity dispersions and X-ray luminosity estimations give the cluster mass as $M_{\rm vir} = 10^{13.73}\,{\rm M}_\odot$ and the estimate given by mass richness is significantly higher $M_{\rm vir} = 10^{14.6}\,{\rm M}_\odot$, whilst we used the published average the lower cluster mass excluding the richness estimate is in as-good or better agreement with model results.

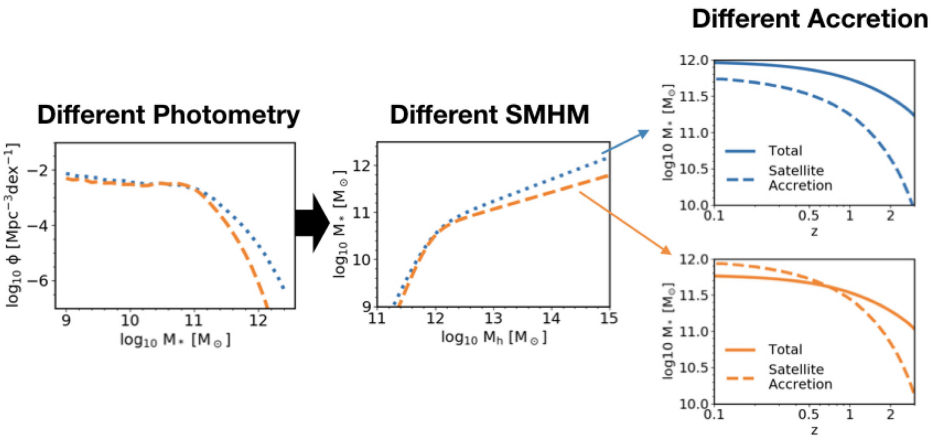


Figure 1. An image showing the steps we follow to connect the differences found in the stellar mass function (left) and the changes in the SMHM relation (SMHM, middle) that propagate into changes in the accretion histories (right). In the right-hand panel dashed lines are mass from satellite accretion and solid lines are total galaxy mass growth. Flatter SMHM relations imply a weaker growth of stellar mass in the central which can be easily overcome by the substantial cumulative growth of merging satellites, rendering the model internally inconsistent.

mass accretion associated with a given central halo mass history. It follows that the galaxy grown using the steeper relation from the enhanced SMF induces more galaxy growth. A flatter high-mass slope induces less growth in the limit where the high-mass slope is flat the central galaxy would not grow with increasing halo mass.

The dashed lines in the rightmost panels are created by calculating the average accretion on to a central galaxy. The majority of accretion is from galaxy mergers that have a low-mass ratio between the central and satellite galaxy. Note, the SMFs show little difference in number density for smaller galaxies (below $M_* = 10^{11}$ M$_\odot$) and thus the low-mass slope of the SMHM relation is unchanged. It follows that the smaller galaxies in a given dark matter assembly history are unchanged and the accretion is similar for both centrals. For the bottom right hand panel showing the accretion history and growth history for a galaxy using the lower SMF it is found that the accretion exceeds the galaxy growth (in this image this is accentuated for clarity). Whereas, for the enhanced SMHM relation the accretion is below that of the galaxy growth. The satellite galaxy accretion history and the central galaxy growth history in a given cosmology are determined by the SMHM relationship and the dark matter halo assembly. In this paper we describe a method that, in a given cosmology, can exclude a set of SMFs over multiple redshift epochs. The evolution of these SMFs is analogous with the growth of the total stellar mass in the Universe over cosmic time. The consistency of the galaxy growth histories and the satellite accretion histories is checked by generating and comparing the ratio of satellite accretion and total mass growth. If the total accretion or rate of accretion is greater than that of the central galaxy mass or galaxy growth rate the set of SMFs is incompatible with the specific ΛCDM cosmology.

The image in Fig. 2 shows a simplification of the processes we follow to derive the SFR by following galaxy populations along their halo mass histories. The plot labelled 1 (green) is the input SMF. The box in red is the statistical dark matter accretion history described in Section 3.1, including the HMF (2a), the central growth histories (2b), and the halo substructure (2c) shown here as a discrete merger tree for visualization purposes. Using the abundance matching routines described in Section 3.3, the SMF (1) and the HMF (2a) are used to create the SMHM relationship (3, black). In Paper I we showed how the dark matter accretion histories (2) and abundance matching (3) can be used to generate distributions of satellites for any central halo. In this work we generate satellite distributions for central haloes at multiple redshifts (4) then test them against simulations and observations in Section 4.2. For each central halo mass track (2b) the average number density of satellites that reach the centre of the halo and merge with the central galaxy, is calculated thus generating the average satellite accretion history (dashed line, 5). Using the central halo growth histories (2b) and the SMHM relation (3), we can generate the average central galaxy growth history (solid line, 5). These two quantities can be compared to check for self-consistency, as described above and shown in Fig. 1. Where a self-consistent central growth and accretion history is found, any deficit between the accreted mass and the growth history is attributed to SFR (delta, 5). The derived SFR for central galaxies (solid line, 6) is compared to observational data (points, 6). Where the SFR prediction generated from the model is found to be consistent with the observed SFR this is a good indication that the model is correct. Additional observational constraints not shown in Fig. 2 can be added to improve the analysis, such as the specific SFR (sSFR) distribution which is discussed in Section 4.4. In future work other constraints such as the pair fraction of galaxies and the intracluster light generated from dynamical process during satellite accretion will also be considered.

3.1 Statistical merging history

Common modelling techniques, such as hydrodynamical, semi-analytic, and traditional semi-empirical models rely on a discrete set of haloes within a simulation 'volume'. These discrete haloes come

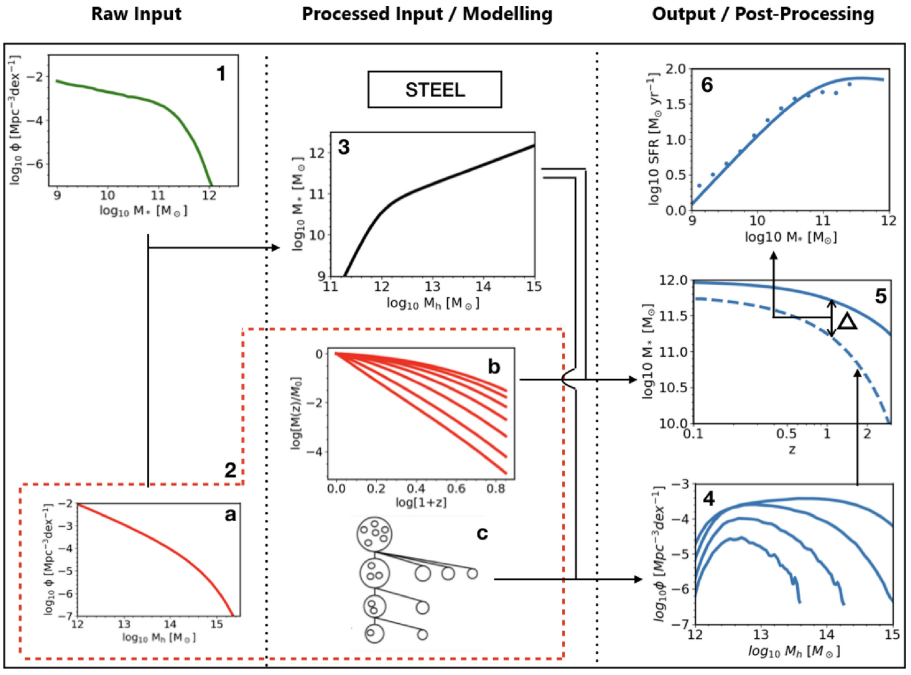


Figure 2. An image showing the constituent steps of the method to generate SFRs in this paper. In brief, the three columns from left to right are raw inputs, derived inputs/modelling, and output/post-processing. The subplots are: (1) The SMF, (2a) The HMF, (2b) Halo mass growth histories, (2c) Accretion histories/Merger tree, (3) The SMHM relation, (4) Group/Cluster satellite richness, (5) Central growth histories/satellite accretion histories, (6) SFR. The SFRs are derived from the difference between the total growth in stellar mass and that from satellite accretion (panel 5).

in three forms: an *N*-body cosmological box, post-processed merger trees, or catalogue of haloes.[3] Due to the inevitably restrained simulated cosmic volumes, all the aforementioned models are biased towards smaller haloes and galaxies, largely missing a statistically comprehensive description of the most massive central and satellite galaxies.

STEEL is instead designed to model all haloes (and subhaloes) within the simulation range without volume or resolution constraints. We remove the dependence on discrete halo sets through the use of a *statistical* dark matter backbone, and simulate all haloes and all mass-ratio mergers[4] with equal weights regardless of their number density. The following steps represent the core method for creating the statistical dark matter accretion histories.

(i) At the redshift of interest $\bar{z}$ we start from the HMF to compute the abundances of (parent/central) haloes for any given central halo bin $[M_{\rm h,cent}(z),\, M_{\rm h,cent}(z) + {\rm d}M_{\rm h,cent}(z)]$.

(ii) Each parent/central halo mass is then followed backwards in time following its average mass growth history, $< M_{\rm halo}(\bar{z}) >$, calculated using the routines from van den Bosch et al. (2014).

(iii) An unevolved subhalo mass function[5] (Jiang & van den Bosch 2016) is then assigned to each central halo mass bin $[M_{\rm h,cent}(z),\, M_{\rm h,cent}(z) + {\rm d}M_{\rm h,cent}(z)]$ at each redshift epoch.

(iv) At each time-step we then calculate the difference in subhalo population between z and $z + {\rm d}z$ to estimate the average number density and masses of subhaloes accreted on to the main progenitor in the redshift interval dz, which we call the 'accreted' subhalo mass function.

(v) Each bin $[M_{\rm h,sub},\, M_{\rm h,sub} + {\rm d}M_{\rm h,sub}]$ of the accreted subhalo mass function is then assigned a dynamical time given the central halo mass bin $[M_{\rm h,cent}(z),\, M_{\rm h,cent}(z) + {\rm d}M_{\rm h,cent}(z)]$ it corresponds to.

(vi) At each redshift epoch we then sum the number densities of subhalo bins $[M_{\rm h,sub},\, M_{\rm h,sub} + {\rm d}M_{\rm h,sub}]$ on each central halo mass

[3]We note that in many cases merger trees and halo catalogues are extracted from the cosmological box of a dark matter only *N*-body simulation. The alternatives being Press–Schecter analytically derived merger trees (e.g. Press & Schechter 1974; Parkinson, Cole & Helly 2008) and halo catalogues obtained by sampling the HMF.

[4]Within the simulation mass range, however, this can be set arbitrarily wide as long as the choice of HMF and empirical techniques are valid in the proposed mass range.

[5]The unevolved subhalo mass function gives the total number density of subhaloes of each mass accreted on to a given central halo over its entire growth history.

track that have not exceeded their dynamical time to create the surviving subhalo mass function.

(vii) Given the infall redshift, mass, and number densities of each subhalo mass bin $[M_{\rm h,sub}, M_{\rm h,sub} + {\rm d}M_{\rm h,sub}]$ we can convolve the resulting satellite halo distribution with the SMHM relation to create the observed distribution of satellites at any epoch. In some model variants we also include additional physical processes to account for the late evolution of satellites after infall (see Paper I).

In this work we also track the number densities of subhaloes that have reached the end of their dynamical time at each epoch. At the time a subhalo reaches its dynamical time the associated satellite galaxy 'merges'[6] with the central galaxy, during these mergers we inject a fraction (40 per cent) of the satellite mass to the intracluster medium (Moster et al. 2018).

3.2 Satellite quenching

The satellite quenching model in STEEL, presented in Paper I, has two components:

(i) A delayed-then-rapid quenching model (Wetzel et al. 2013), according to which satellite galaxies upon entering a halo continue to form stars as if they were on the star formation main sequence until their quenching time-scale, $\tau_{\rm q}$, has elapsed. After a time $\tau_{\rm q}$ the SFR of the satellites is rapidly quenched over the fading time-scale $\tau_{\rm f}$.[7]

(ii) The second component used is the halo mass-dependant cut-off (Fillingham et al. 2018), which envisions that satellite galaxies below a given stellar mass (dependant on host halo mass) are immediately quenched.

The delayed-then-rapid quenching model is improved using the latest dynamical quenching results from Cowley et al. (2019) updating the model presented in Paper I to include a redshift dependence in all quenching time-scales

$$\tau_{{\rm q},z} = \tau_{\rm q} * (1 + z_{\rm infall})^{-3/2},$$
$$\tau_{{\rm f},z} = \tau_{\rm f} * (1 + z_{\rm infall})^{-3/2}. \quad (1)$$

Additionally, we include a pre-processed fraction, inspired by the results of Wetzel et al. (2015), implementing a mass-dependent fraction such that a minimum of 30 per cent and a maximum of 60 per cent of galaxies are pre-possessed with a transitional mass range between 10^6 and 10^8 M$_\odot$. The SFR of a satellite galaxy after infall is then given by

$${\rm SFR}(t, t_{\rm infall}, M_{*,\rm infall}) = {\rm SFR}_{t_{\rm infall}} * \begin{cases} 1, & t > t_{\rm q} \\ {\rm e}^{\left[-\frac{t_{\rm q}-t}{\tau_{\rm f}}\right]}, & t < t_{\rm q}. \end{cases} \quad (2)$$

3.3 Abundance matching

In this work we populate dark matter haloes with galaxies using the abundance matching technique where galaxies are assigned to haloes by comparing the relative abundances of galaxies and haloes. For the abundance matching we use the central haloes from the HMF described in Section 2.1, and a subhalo mass function subdivided by redshift of infall generated from STEEL. Subhaloes are assumed to follow the central SMHM relation at infall. We simplify our abundance matching by using a frozen model such that baryonic evolution after infall (stripping, star formation, etc.) is not included. The latter assumption provides a good approximation as after infall the dominant factor determining the abundances of satellite galaxies is the dynamical time and not evolutionary processes (Paper I).

To fit SMFs over multiple epochs we convolve our HMFs with a parametric SMHM relation similar to that proposed by Moster et al. (2010),

$$\begin{aligned} M_*(M_{\rm h}, z) &= 2M_{\rm h}N(z)\left[\left(\frac{M_{\rm h}}{M_{\rm n}(z)}\right)^{-\beta(z)} + \left(\frac{M_{\rm h}}{M_{\rm n}(z)}\right)^{\gamma(z)}\right]^{-1} \\ N(z) &= N_{0.1} + N_z\left(\frac{z-0.1}{z+1}\right) \\ M_{\rm n}(z) &= M_{\rm n,0.1} + M_{{\rm n},z}\left(\frac{z-0.1}{z+1}\right) \\ \beta(z) &= \beta_{0.1} + \beta_z\left(\frac{z-0.1}{z+1}\right) \\ \gamma(z) &= \gamma_{0.1} + \gamma_z\left(\frac{z-0.1}{z+1}\right). \end{aligned} \quad (3)$$

In what follows we adopt both the cmodel and PYMORPH SMF described in Section 2.2 at redshift $z = 0$ to constrain the parameters N, M, β, and γ (normalization, knee, low-mass slope, and high-mass slope). We use only the central SMF, using the Yang et al. (2012) central/satellite identification, and central HMF. The fit is performed using a Markov Chain Monte Carlo (MCMC), implemented using the PYTHON package EMCEE (Foreman-Mackey et al. 2013), over a large parameter space ($P_{M,N,\beta,\gamma}$) covering all four parameters. Given a point in parameter space $P_{M_i,N_i,\beta_i,\gamma_i}$, the SMF is constructed using the HMF and the SMHM relation. Each bin of parent halo mass is associated with a Gaussian distribution of stellar mass with scatter 0.15 dex. This distribution is multiplied by the halo mass number density to convert to galaxy number density which are added to the relevant stellar mass bins of the SMF in construction. This operation is then repeated over all mass bins of the HMF to produce the complete central SMF. For each point $P_{M_i,N_i,\beta_i,\gamma_i}$ in the parameter space, the SMF associated with that point is compared via a likelihood function to the observed SMF to provide the MCMC with the probability that the given point is the 'true' SMHM relationship.

We then fit to the Davidzon et al. (2017) data both uncorrected and corrected for the cmodel and PYMORPH fits, respectively (see Section 2.2 for details). At high redshift we use the central and subhalo mass functions initializing satellites at infall as described above.[8] For central haloes the method is the same as detailed above, however, as we use the total SMFs at high redshift we also include the total unevolved surviving subhalo mass function in the abundance matching. We assume that a halo before infall hosts a central galaxy; under this assumption we use the central SMHM relation to assign satellite galaxy stellar mass at the point of accretion. For the latter we must have information about the redshift of infall for subhaloes. We obtain from STEEL the unevolved

[6] In a statistical model satellite galaxies are not strictly merging as there is no central galaxy to merge with, instead we collect statistics of merging satellites at each epoch. Using post-processing techniques we use these statistics obtain information on the average merging history of central galaxies.

[7] The absolute quenching time is given by $t_{\rm q}(= t_{\rm infall} - \tau_{\rm q})$.

[8] Ideally, as for low redshift, we would use the centrals only as we are primarily concerned with the central SMHM relation. However, lacking a well-defined central SMF at high redshift, this method represents a reliable way to extend the model to higher redshifts.

Table 1. The abundance matching results for the cmodel and PYMORPH data. The errors are the 16th and 86th percentile from the MCMC fitting.

	M_n	N	β	γ	$M_{n,z}$	N_z	β_z	γ_z
cmodel	$11.91^{+0.40}_{-0.34}$	$0.029^{+0.018}_{-0.013}$	$2.09^{+1.21}_{-1.02}$	$0.64^{+0.11}_{-0.10}$	$0.52^{+0.24}_{-0.19}$	$-0.018^{+0.005}_{-0.004}$	$-1.03^{+0.049}_{-0.34}$	$0.084^{+0.20}_{-0.14}$
PYMORPH	$11.92^{+0.39}_{-0.36}$	$0.032^{+0.016}_{-0.012}$	$1.64^{+0.85}_{-0.73}$	$0.53^{+0.11}_{-0.11}$	$0.58^{+0.15}_{0.19}$	$-0.014^{+0.007}_{-0.006}$	$-0.69^{+0.29}_{-0.36}$	$0.03^{+0.154}_{-0.147}$

surviving subhalo mass function as contributed by each redshift of infall. Each contributing part is calculated using the SMHM relation at the redshift of infall and added to the central SMF using the same method as with the centrals. The total SMF is compared, at each redshift step available, to the data via the likelihood function to give the probability that the given point is the 'true' evolution parameters. The abundance matching best-fitting parameters and associated errors for both the cmodel and PYMORPH are given in Table 1, and plots showing the cross-sections of the parameter space are shown in Appendix A.

In Fig. 3 we show the results of our abundance matching to the PYMORPH and cmodel central SMFs. The PYMORPH fit is steeper above the knee compared to either the cmodel or the Moster et al. (2013) model fits, as expected given the larger number density of massive galaxies found applying the Sérsic–exponential model (e.g. Shankar et al. 2014; Kravtsov, Vikhlinin & Meshcheryakov 2018). The low-mass slope for both PYMORPH and cmodel are almost identical as the galaxies in this range are not affected by the photometric choice. Differences between the fits from this work and Moster et al. (2013) are due to our selection of using only central haloes/galaxies as opposed to the total population, and the SMFs shown in the right-hand panel are lower than even cmodel are therefore missing massive galaxies.

4 RESULTS

4.1 Halo merger rates

STEEL implements a statistical dark matter merging history, thus as a very first step we check STEEL's performance on reproducing halo merger rates as extracted from *N*-body dark matter-only simulations. We explore the evolution of the merger rate of haloes with a mass ratio greater than $f = M_{\rm h,sat}/M_{\rm h,cen}$. The merger rate is calculated from STEEL by integrating the 'unevolved subhalo mass function accretion' (δUSHMF) above the mass ratio limit

$$\frac{{\rm d}N}{{\rm d}z}\left(M_{\rm h,cen}\right) = \int_{M_{\rm h,cen}*f}^{\infty} \delta \quad {\rm USHMF}([z, z+\delta z], M_{\rm h,cen}, M_{\rm h,sat}){\rm d}M_{\rm h,sat}. \tag{4}$$

In Fig. 4 the merger rate from STEEL, shown by lines, is in good agreement with the best-fitting merger rate relations from the Millennium simulation given by Fakhouri, Ma & Boylan-Kolchin (2010), shown as shaded regions.[9] The slight deviation at low redshift derives from the predicted growth history of our input potential wells given by van den Bosch et al. (2014), a lower mass growth rate leads to a lower accretion rate. This deviation is due to differences in the algorithms used to link haloes between simulation outputs and build merger trees used by Fakhouri et al. (2010) and van den Bosch et al. (2014)

[9]It should be taken into consideration that the results from STEEL presented here and the fits from Fakhouri et al. (2010) are in different cosmologies. We show STEEL halo accretion using the Millennium cosmology used in the aforementioned work in Appendix D.

4.2 High-redshift clusters

We here extend the group and cluster satellite richness analysis from Paper I to high redshift. In Paper I it was found that dynamical friction and, to a second order, abundance matching, are the dominant factors in the distribution of satellite galaxies in groups and clusters above $M_{*,\rm sat} > 10^{10}\,{\rm M}_{\odot}$. In this section, for a more rounded view of the satellite galaxy population, we display the results for the full STEEL model which includes star formation, dynamical quenching, and stripping to evolve satellites after infall. The latter effects, despite being of lower order than dynamical friction or abundance matching, are included to be able to compare to data other than cluster richness, such as the satellite sSFR distribution.

Fig. 5 shows the satellite number density per halo mass bin. For each central halo mass the cosmic number density, similarly to the number density presented in the cumulative SMFs, is calculated for satellites above a mass threshold for each halo mass bin. The predicted halo richness from STEEL, using the PYMORPH SMHM relation, is shown in this plot as solid lines. Low-redshift SDSS data are shown as a grey band, cluster data detailed in Section 2.3 are open symbols. The predictions from the Illustris TNG100 simulation (Nelson et al. 2018; Springel et al. 2018) are shown with crosses. The markers and lines in the figure are colour coded based on redshift, as indicated by the colour bar on the right.

In Fig. 5 we see that at higher redshift there are fewer massive satellites overall. This reduction is caused by several contributing effects. First, at high redshift there is an absence of high-mass haloes, which have not had time to form. Due to the lack of hosts at higher masses, the right hand side of the distribution tightens. Secondly, at high redshift the HMF is lower at any given mass, causing a shift of the satellite host halo distributions towards lower masses. Finally, as the process of formation and merging takes several gigayears to complete massive satellites are found less frequently at high redshift, and thus the number densities of the most massive satellites reduce faster than the lower mass ones.

We show STEEL is a good match to the Wang et al. (2016) cluster at redshift $z = 2.5$. We also achieve an adequate match to the cluster survey from Wen & Han (2018), especially in the mass range above $\rm M_{sun} > 10^{10.5}\,M_{\odot}$. We also achieve similar results to the Illustris TNG100 simulation, though the TNG100 output is marginally higher at all redshifts. STEEL improves upon TNG100 in terms of the shape and breadth of this distribution. For example, at high mass and redshift STEEL resolves the turnover for satellites above $\rm M_{sun} > 10^{11}\,M_{\odot}$, whereas TNG100 is too limited in volume to cover the high-mass ranges. The limitations in volume prevent TNG100 from simulating clusters such as those presented in Fig. 5 (Wang et al. 2016; Wen & Han 2018). In this respect STEEL becomes an excellent bridge between the capabilities of a high-resolution

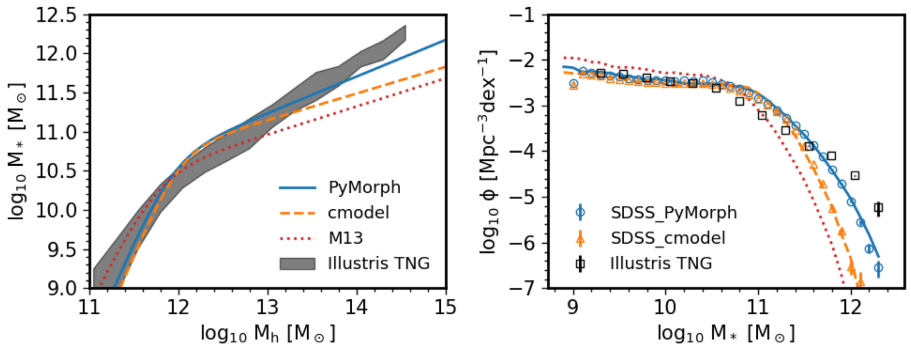


Figure 3. Left: The SMHM relation at redshift $z = 0.1$. The PYMORPH (blue solid line) and cmodel (orange dashed line) fits from this work are both for central haloes/galaxies, the fit from Moster et al. (2013) (hereafter M13, red dotted line) is for all haloes/galaxies. The grey band is the relation from Illustris TNG100. Right: SMFs created using the central HMF and the three SMHM relations compared to PYMORPH (blue circles) and cmodel (orange triangles) central SMFs. The black squares are the SMF from Illustris TNG100.

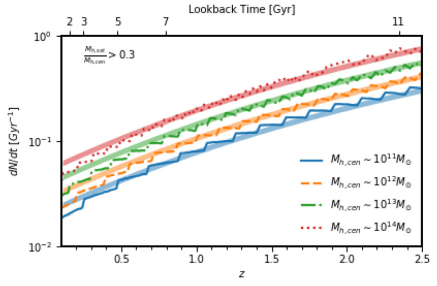


Figure 4. The evolution of merger rate per Gyr at fixed halo mass. Lines are from STEEL, shaded bands are the analytic fits from Fakhouri et al. (2010). Halo masses shown are $M_{\rm h,cen}$: 10^{11}, 10^{12}, 10^{13}, 10^{14} $\mathrm{M}_\odot\,h^{-1}$ as labelled.

hydrodynamical simulation and the massive cluster observations at high redshift.

4.3 Connecting central mass accretion to star formation rate

We calculate from STEEL, using the PYMORPH SMHM relation, the relative contributions to the average stellar mass growth of central galaxies from satellites and star formation history (SFH), as shown in Fig. 6 . For three galaxy mass bins (10^{11}, $10^{11.5}$, 10^{12} $\mathrm{M}_\odot$) selected at $z = 0.1$, the average growth history (total, solid lines) is derived by following the host halo mass track, and the stellar mass track is implied by imposing abundance matching at all redshifts. The stellar mass history assigned by abundance matching, is naturally independent of any galaxy merger modelling assumptions from STEEL. The total accretion from satellites (accretion, dashed lines) is computed from the expected satellite accretion along halo mass tracks. For each galaxy an SFH (dotted lines) may then be calculated. The SFR is tuned such that it provides the correct SFH to account for the difference between the mass growth expected from abundance matching and the cumulative satellite stellar mass accretion.[10] When calculating this difference we also take into account the stellar mass-loss rate (MLR) due to stellar recycling using the relations from Moster et al. (2018),

$$f(\tau_{\rm ml}) = 0.05 \ln\left(\frac{\tau_{\rm ml}}{1.4\,\mathrm{Myr}} + 1\right), \tag{5}$$

$$\mathrm{MLR}(t) = \frac{\sum_{t'=t_{\rm infall}}^{t} \mathrm{SFH}(t')(f[t'-(t-\delta t)] - f[t'-t])}{\delta t}. \tag{6}$$

The SFR–stellar mass relation derived from this method is fitted with a double power law that evolves with redshift (for more details on the fit see Appendix B), as shown in Fig. 7. At lower redshift the normalization decreases, the peak of the distribution shifts to lower masses, and the turnover after the peak is steeper. In Fig. 7 we also show the same three galaxy population tracks from Fig. 6, discussed below, as black lines. These tracks show how the galaxy population evolves in SFR with redshift. The population tracks show a gradual increase in SFR and then a turnover before dropping sharply, as they transition to a satellite accretion-dominated regime. It is found that smaller galaxies grow for longer time-scales with increasing star formation, whilst larger galaxies start with higher SFR and transition to an accretion-dominated phase much earlier in time.

The top row of Fig. 6 shows the total mass of the galaxy and the total contributed by both accretion and the SFH. The middle row shows the fractional contributions from star formation and satellite accretion from $z = 3$. The bottom row shows the instantaneous mass growth from star formation and satellite accretion. There are two definitions we can use to determine the epoch after which a galaxy transitions into a merger-dominated state. First, we could define the 'cumulative transition' as when the galaxy has accreted more mass than it has created from star formation processes

[10]This method directly links the SFR to the accreted mass from satellites. However, in our model satellites to grow in mass after infall (i.e. 'non-frozen' to follow the terminology of Paper I), we therefore recalculate the full satellite accretion on to the central galaxies updating their mass using the new SFR. Using the updated accretion the SFR is recalculated beginning an iterative process. However, this iterative process of recalculation ends after one loop as the rederived accretion is found to be nearly identical, as expected from the results of Paper I.

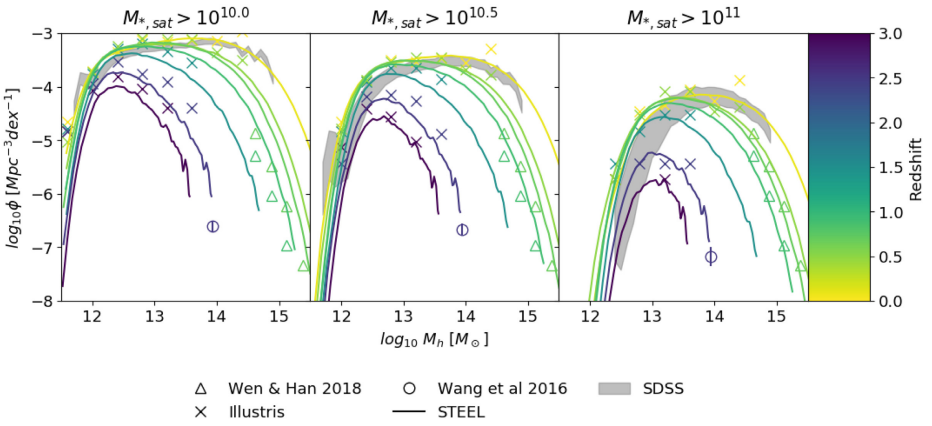


Figure 5. The number density distribution of satellites per parent halo mass predicted from STEEL, using the PYMORPH SMHM relation, at multiple redshift epochs (solid lines). The grey band is the data from SDSS at redshift $z = 0.1$. Also included are the high-redshift cluster data from Wang et al. (2016) (circles) and Wen & Han (2018) (triangles). We also compare to the outputs from the Illustris simulation using the TNG100 data (crosses). Each data point and line are given a colour associated with their redshift (the bar on the right provides the colour coding key).

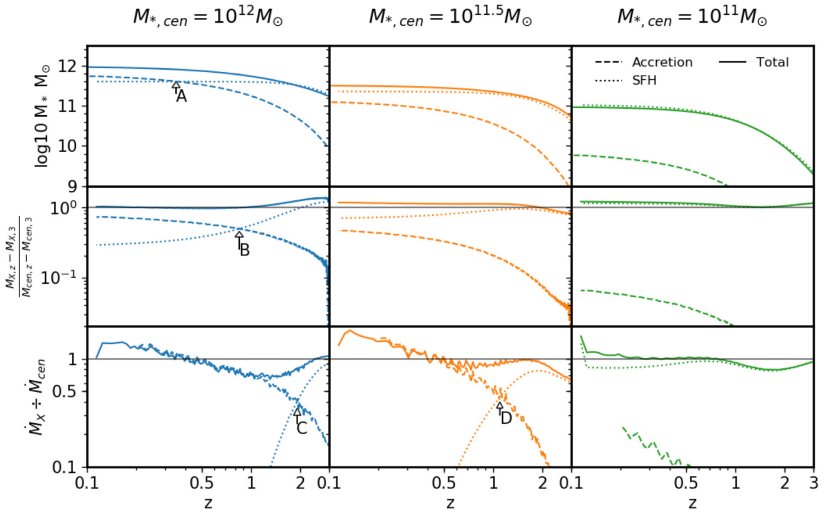


Figure 6. Three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,\mathrm{cen}} = 10^{12}$, $10^{11.5}$, and 10^{11} [$M_\odot$] in blue, orange, and green, respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line, and the mass from star formation shown with a dotted line. The top panels show the total mass of the central (solid lines) and the total mass gained from accretion or star formation. The middle panels show the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panels show the ratio of the mass accretion rate from satellite galaxies, the SFR, and the mass growth rate of the central galaxy predicted by abundance matching. The black horizontal lines in the second and third rows are at unity. The solid lines showing the sum of the other two factors should be close to or on the unity lines. The labels A and B point to where the cumulative mass from accretion overtakes the cumulative mass from star formation. The labels C and D point to where the instantaneous accretion overtakes the SFR.

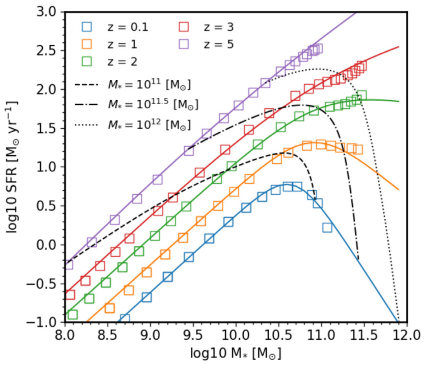


Figure 7. The SFR–stellar mass relation derived from following central galaxy populations along halo mass histories at redshifts $z = 0.1, 1, 2, 3, 5$. The data extracted from the post-processing of STEEL are shown by coloured crosses and the double power-law fits are shown as lines in corresponding colours. The three black lines are the evolution of the galaxy populations selected at redshift $z = 0.1$ with masses $M_* = 10^{11}, 10^{11.5}, 10^{12}(\mathrm{M}_\odot)$ presented in Fig. 6.

(Points A and B). Secondly, we define the 'instantaneous transition' as the epoch when the growth rate from mergers overtakes the growth rate from star formation (Points C and D). More massive galaxies transition earlier to merger dominated growth under both definitions. However, all galaxies transition earlier under the second (instantaneous) definition. The masses shown in Fig. 6 show three cases of relevant galaxy accretion tracks. The $M_*^{z=0} = 10^{12}\,\mathrm{M}_\odot$ galaxy growth curve at low redshift is always dominated by satellite accretion. In the top and middle rows we see that more mass has been accreted than produced by star formation, and in the bottom row we see the accretion rate overtook the SFR at redshift $z = 2$. The $M_*^{z=0} = 10^{11.5}\,\mathrm{M}_\odot$ galaxy growth curve has more mass created from star formation than satellite accretion. However, the galaxy population has a higher rate of accretion rate than SFR since redshift $z = 1$. The final population shown at $M_*^{z=0} = 10^{11}\,\mathrm{M}_\odot$ is star formation dominated under both cumulative and instantaneous definitions. At redshift $z = 0$ we find the transition masses for the total mass ratio and the instantaneous ratio to be at $M_* = 10^{11.7}\,\mathrm{M}_\odot$ and $M_* = 10^{11.1}\,\mathrm{M}_\odot$, respectively.

We show in Fig. 8 the satellite accretion for the cmodel abundance matching using the same template as Fig. 6. In Fig. 8 we obtain a lower limit for the accretion rate by including stripping but not star formation in the satellites thus minimizing their mass through environmental processes. We find for the high-mass galaxies, which are above the knee of the SMHM relation, even the lower limit for the accretion has an instantaneous rate greater than the growth rate of the galaxy as seen in the bottom row. This makes the cmodel SMHM relation used within our dark matter accretion model *non-physical*: steeper SMHM relations, such as the one found with the PYMORPH photometry, are favoured by hierarchical assembly. We recall, as explained in Fig. 1, that too flat SMHM relations introduce global stellar mass growth histories that are even lower than what is expected from total satellite accretion rendering the models internally inconsistent. For completeness we also tested a range of dynamical time options. In all cases, even when the merging time is increased by a factor of two,[11] the accretion rate exceeds the growth rate and cmodel photometries can be excluded. Further to this we also tested variations on the mass-loss in mergers and the amount of mass lost to tidal stripping of the satellites; in the case where the tidal is doubled with respect to the reference model and the mass-loss during a merger is set to 60 per cent up from 40 per cent, the cmodel photometries remain internally inconsistent. We are confident that under all circumstances cmodel photometries can be considered internally inconsistent.

4.4 Specific star formation rate distribution

Fig. 9 shows the sSFR distribution of satellites in three mass ranges, as labelled, chosen to probe transitions found in observational data (Bernardi et al. 2011, 2014; Cappellari et al. 2013). The solid blue line and the dashed black lines show the satellite and central sSFR from STEEL, respectively, while the grey histogram shows the satellites from SDSS and the unfilled histogram shows the centrals in SDSS.

STEEL accurately captures the key trends in the distributions, such as bimodality, which is seen in both the central and satellite populations. The central population below $M_* = 10^{10.5}$ ($\mathrm{M}_\odot$) is mostly star forming whereas the satellites show signs of quenching. In the intermediate-mass range a fraction of the centrals become quenched and the satellites show a strong quenching effect. In the highest mass range all galaxies show strong quenching features with little star formation. Whilst still not an exact match to the SDSS distribution, we find that including a redshift dependence in the dynamical quenching provides a better fit than the model used in Paper I. The central sSFR is calculated using the SFR presented in Section 4.3, which use the PYMORPH SMHM relation. Each central mass is assigned an SFR with a scatter of 0.2 dex. To account for the fraction of galaxies that are quenched via mergers at each stellar mass we modify the assigned SFRs by setting a fraction of galaxies equal to the elliptical fraction from Section 5.3 to have an sSFR of 10^{-12} (yr^{-1}) with a scatter of 0.2 dex and in turn increase the SFR of the remaining galaxies to maintain the same average SFR for the population. This approach tests if mergers alone can account for the bimodality found in the central sSFR, the high-mass centrals $>10^{11.3}$ ($\mathrm{M}_\odot$), but produces an inadequate fit to the SDSS centrals at masses lower than $10^{10.5}$ ($\mathrm{M}_\odot$). The discrepancies in the location of the star-forming population are likely caused by the imperfect fit to observed SFR as seen in Fig. 10 and the deficit of quenched galaxies in the lower mass cuts is likely due to causes of quenching that are not merger related (e.g. AGN feedback).

5 DISCUSSION

Using STEEL we have presented a consistent picture of group and cluster richness across several orders of magnitude in mass, and satellite accretion histories over 11 Gyr of the Universe's history. It is essential for a model that aims to predict the hierarchical growth of galaxies, that both the central and satellite SMFs are well reproduced at all redshifts. STEEL uses state-of-the-art statistical accretion histories and powerful abundance matching techniques to ensure we have the essential consistency with observed galaxy number densities by design.

[11] Dynamical time factors higher than 2 are shown to not reproduce the satellite distributions in Paper I

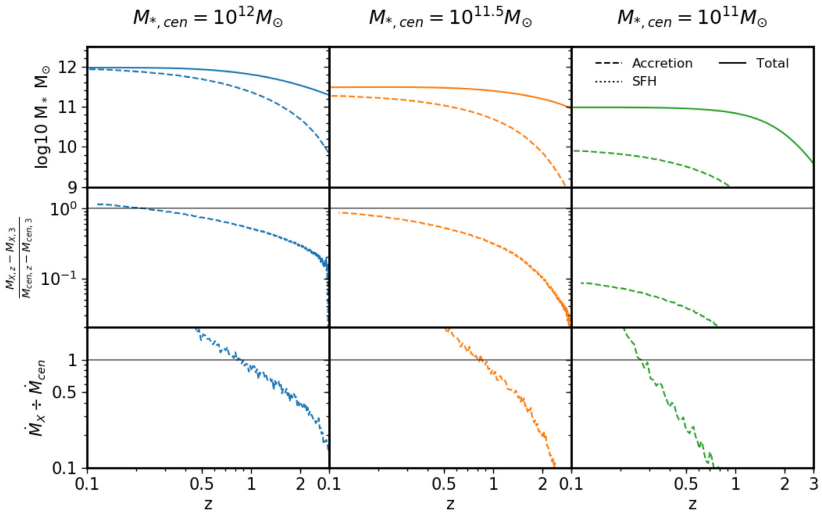


Figure 8. Same format as Fig. 6 but for the cmodel photometry. It is clear that this model is internally non-physical as the accretion via satellites (dashed lines) rapidly overshoots the total growth in stellar mass (solid lines) implied by the underlying growth host halo growth, as evident in the middle and bottom rows.

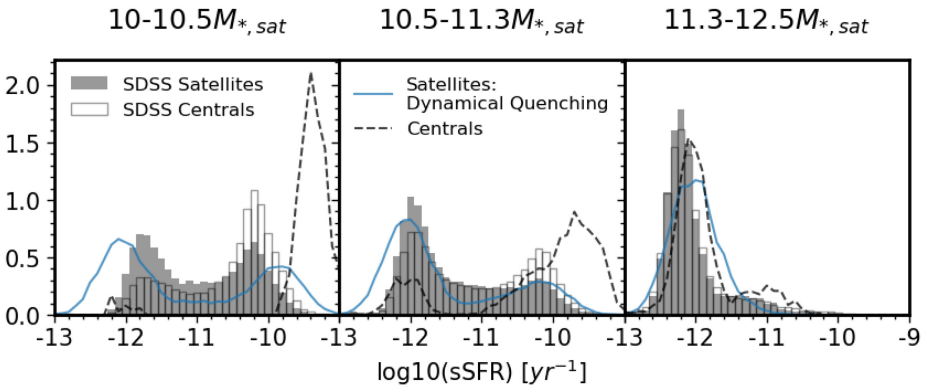


Figure 9. We show the sSFR of satellites and centrals compared to SDSS in three mass bins selected to mirror proposed breaks in the galaxy main sequence. The SDSS data for satellites and centrals are filled and unfilled histograms, respectively. The STEEL result for the satellites is the solid blue line and the post-processed central result is the dashed black line.

5.1 High-redshift clusters

Galaxy groups and clusters represent an excellent laboratory to test theories on galaxy evolution. The rich cluster environments are observable up to high redshift and contain some of the most massive galaxies. Exploring the richness of the environments around massive galaxies provides an excellent constraint to hierarchical assembly predicted by ΛCDM cosmology at the most extreme masses (Shankar et al. 2015). In this work we show a singular cluster reported in Wang et al. (2016). Other models (e.g. Henriques et al. 2015) have been unable to reconcile this cluster within a ΛCDM framework. We found in Fig. 5 that STEEL is able to predict the existence of such massive objects. However we concur with Wang et al. (2016) that these objects are rare and their absence in traditional simulations could be simply attributed to poor statistics and not necessarily to the implied physical model. With large-scale

surveys such as EUCLID coming online, a well-tuned statistical model could more easily place robust constraints on high-redshift cluster formation.

5.2 Central assembly

In Section 4.3 we found one of the major factors in regulating the *in situ* and *ex situ* accretion pathways to be the *shape* of the SMHM relation. A shallower low-mass slope causes larger amounts of satellite accretion as smaller haloes, with much higher number density, are initialized with larger satellite galaxies. Similarly to Shankar et al. (2006) and Moster et al. (2018), we find the high-mass slope to undergo only a small amount of evolution with increasing redshift, this implies the growth of central galaxies is directly linked to the steepness of the high-mass slope and the growth of the host halo. The flatter the high-mass slope of the SMHM relation, the less growth is expected in stellar mass following the assembly of the host dark matter halo. In turn, a weak evolution in the stellar mass content of the central galaxy can be in tension with what is expected from satellite accretion, especially for the most massive galaxies. We discussed that the slope of the high-mass end of the SMF and implied slope of the SMHM relation strongly depend on the choice of light profile, background subtraction, and mass-to-light ratios. However, not all resulting SMFs provide physically self-consistent results in a ΛCDM Universe. Steeper SMHM relations, such as those predicted by PYMORPH-based SMFs (Bernardi et al. 2013), produce more consistent central and satellite accretion stellar mass growths. In addition to models with different SMHM slopes, we also tested models with the dynamical time varied by ±20 per cent, within the range of possible dynamical times predicted in Paper I constrained by satellite richness. This relatively modest alteration has a minor effect on the satellite accretion rate and mass contribution to the central. In this work we find the transitional stellar mass, above which dry mergers progressively become the major contributor to galaxy growth, to be $M_* = 10^{11.1}$, see Fig. 6. The latter is consistent with previous findings (e.g. Bernardi et al. 2011; Cappellari 2013; Shankar et al. 2013).

By following the statistical dark matter accretion histories we were able to use the central mass tracks and abundance matching to obtain a growth history for central galaxies. Subtracting from the latter at each time-step the cumulative stellar mass from satellite accretion, we created a 'SFR' interpreted as the remaining mass required to build the central mass. Our methodology is similar to the continuity approach based on Leja et al. (2015) used in Paper I, but with the key difference that here we follow halo growth instead of galaxy number density. The resulting SFR for galaxies is notably different to that of Tomczak et al. (2014), used in Paper I. At all redshifts the turnover is notably different, with SFR for masses above the turnover decreasing sharply at low redshift. For masses below the turnover, at $z < 1$ the SFR is lower by 0.3 dex, and at $z > 1$ the SFR is higher by 0.1–0.2 dex. Recent work, where the SFHs are properly accounted for when measuring SFRs, has suggested that the previous determinations of SFRs using UV+IR are 0.1–1 dex too high (Leja et al. 2019) and cannot be reconciled with the growth of the SMF (Leja et al. 2015; Lapi et al. 2017). Our SFR is consistent with the results of Leja et al. (2019), as reported in Fig. 10. The excellent match to Leja et. al.'s independent estimates further supports the idea that a more robust method to derive more reliable SFRs is to follow galaxy assembly along host halo growth histories (see e.g. Moster et al. 2018).

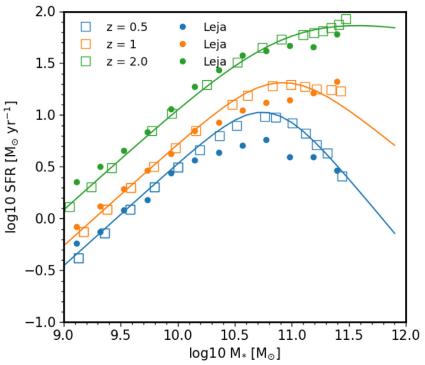


Figure 10. We show the SFR–stellar mass relationship from Fig. 7 at redshifts $z = 0.5$, 1, 2 (blue, orange, and green, respectively, STEEL data are crosses and fits are solid lines). In this plot we compare with the observed SFR from Leja et al. (2019) shown as filled circles with corresponding colours denoting corresponding redshift.

5.3 Central morphologies

Mergers are thought to be one of the drivers for morphological transformation, size growth, and other galaxy changes (Bournaud, Jog & Combes 2007; Hopkins et al. 2009, 2010a; Shankar et al. 2011; Fontanot et al. 2015). Broadly inspired by the results of hydrodynamic simulations, a number of analytically based cosmological models have generally assumed that major mergers, in particular, with a mass ratio of at least $M_{\rm sat}/M_{\rm cen} > 0.25$, are effective in destroying discs and in forming ellipticals (Baugh 2006; Malbon et al. 2007; Bower et al. 2010). Given the very promising results of STEEL in predicting satellite number densities in different environments and epochs, we here take a step further and explore whether STEEL's cumulative number of major mergers is able to account for the local fraction of elliptical galaxies. For each central mass track we evolve the fraction of galaxies that have had a merger with stellar mass ratio greater than 0.25 since redshift $z = 3$. Fig. 11 shows the probability/fraction of central galaxies that have experienced a merger above the mass threshold of 0.3 at redshifts $z = 0.1, 0.65, 1.75$, while the black triangles show the T-Type-selected elliptical fraction from the SDSS catalogue. We find that applying this simple recipe to the merging number densities from STEEL creates a good match to the elliptical fraction in the local universe.

Despite the noticeably good agreement between model predictions and data in Fig. 11, we stress that different input SMHM relations can, as proven in this work, substantially affect the accretion rate which in turn will modify the number of galaxies experiencing major mergers. It follows that any cosmological galaxy evolution model that uses mergers as a physical driver for galaxy transformation should first simultaneously and self-consistently closely reproduce SMFs, the SMHM relation, and satellite distributions at high redshift.

6 CONCLUSIONS

In this second paper on our STatistical sEmi-Empirical modeL, STEEL, we proved that STEEL can successfully reproduce galaxy satellite richness also at high redshifts. Its innovative design,

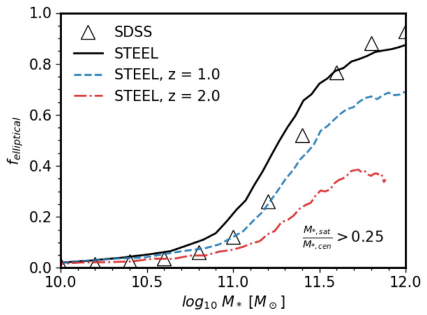


Figure 11. We show at three redshift steps the predicted fraction of ellipticals as a function of stellar mass. The lines are the predictions from STEEL and the tringles are the T-Type selected elliptical fraction from SDSS at redshift $z = 0.1$.

unconstrained by volume or mass resolution, allows STEEL to predict the number densities of even the rarest objects in the Universe at the highest redshifts, a fundamental test for dark matter and galaxy evolution theories though currently inaccessible by cosmological hydrodynamic simulations.

Given the success of STEEL in reproducing satellite richness at different cosmic epochs and environments, we can in turn predict more reliable galaxy merger rates, using central growth rates implied by the central mass track of our statistical dark matter accretion histories and abundance matching. We found that SMHM relations with shallow high-mass slopes create central growth histories that are physically inconsistent with the expected satellite merger rate. We found that steeper SMHM relations at the high-mass end, as induced by the latest determinations of the SMFs based on Sersic–exponential photometry, are favoured against shallower SMHM relations, based on outdated determinations of the SMF. The total stellar mass growth of a galaxy is mostly due to satellite mergers and/or star formation. A flatter SMHM relation, however, naturally implies, for a given increase in host halo mass, a much weaker growth in the stellar mass of the central galaxy than in the case of a steep SMHM relation. The accretion via satellites could then be substantial enough to overshoot the moderate growth in the central galaxy rendering the model internally physically inconsistent.

By safely assuming the difference in central growth rate ($\dot{M}$) and satellite accretion rate is attributable to the star formation in the central galaxy, we predict SFHs and an SFR–stellar mass relations. The latter approach is qualitatively similar to a continuity equation (e.g. Leja et al. 2019, and Paper I), but more accurate as it is developed along the accretion tracks of host haloes so better follows galaxy populations. We find our resulting SFRs to be in excellent agreement with the latest cutting-edge observational measurements by Leja et al. (2019), based on multiparameter Bayesian analysis.

Finally, following traditional models of galaxy evolution, we use our improved galaxy merger rate to predict the fraction of central galaxies as a function of mass that have been transformed into ellipticals via major mergers, (where the stellar mass ratio of the central to satellite is greater than 1/3). We find this fraction to be in excellent agreement with centrals selected as ellipticals via T-Type in SDSS. We use this elliptical fraction and our derived SFR to create a distribution of sSFRs. We find this basic and common assumption to form ellipticals in analytic cosmological models to be sufficient to also reproduce the bimodality in SFRs of massive galaxies above $M_* > 10^{11.3}$ $M_\odot$. Below this stellar mass threshold, we find a too high fraction of star-forming galaxies, which implies additional quenching mechanisms, beside major mergers, must be included in the models.

Our results are of the utmost importance to predict robust and self-consistent SMHM relations and to generate reliable mock catalogues for the next generation of extragalactic surveys such as Euclid and LSST.

ACKNOWLEDGEMENTS

We acknowledge extensive use of the Python libraries ASTROPY, MATPLOTLIB, NUMPY, PANDAS, and SCIPY. PJG acknowledges support from the STFC for funding this PhD. FS acknowledges partial support from a Leverhume Trust Research Fellowship and from the European Union's Horizon 2020 programme under the AHEAD project (grant agreement #654215).

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abolfathi B. et al., 2017, ApJS, 235, 42
Asquith R. et al., 2018, MNRAS, 480, 1197
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Bernardi M., Roche N., Shankar F., Sheth R. K., 2011, MNRAS, 412, L6
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS, 443, 874
Bernardi M., Meert A., Sheth R. K., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016, MNRAS, 455, 4122
Bernardi M., Fischer J.-L., Sheth R. K., Meert A., Huertas-Company M., Shankar F., Vikram V., 2017, MNRAS, 468, 2569
Bournaud F., Jog J., Combes F., 2007, A&A, 476, 1179
Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, MNRAS, 407, 2017
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buchan S., Shankar F., 2016, MNRAS, 462, 2001
Cappellari M., 2013, ApJ, 778, L2
Cappellari M. et al., 2013, MNRAS, 432, 1862
Cattaneo A., Mamon G. A., Warnick K., Knebe A., 2011, A&A, 533, A5
Chabrier G., 2003, PASP, 115, 763
Cowley W. I. et al., 2019, ApJ, 874, 114
Davidzon I. et al., 2017, A&A, 605, A70
de Vaucouleurs G., 1948, Ann. Astrophys., 11, 247
Dekel A. et al., 2009, Nature, 457, 451
Despali G., Giocoli C., Angulo R. E., Tormen G., Sheth R. K., Baso G., Moscardini L., 2016, MNRAS, 456, 2486
De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, MNRAS, 366, 499
Diemer B., 2017, Technical report, COLOSSUS: A python toolkit for cosmology, large-scale structure, and dark matter haloes. Harvard-Smithsonian Center for Astrophysics, http://arxiv.org/abs/1712.04512
Fakhouri O., Ma C.-P., 2010, MNRAS, 401, 2245
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 2278, 2267
Fillingham S. P., Cooper M. C., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., Wheeler C., 2018, MNRAS, 477, 4491
Fontanot F., Macciò A. V., Hirschmann M., De Lucia G., Kannan R., Somerville R. S., Wilman D., 2015, MNRAS, 451, 2968
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Technical report, emcee: The MCMC Hammer. http://dan.iel.fm/emceeunderthe MITLicense

González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, ApJ, 735, L34
Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580
Groenewald D. N., Skelton R. E., Gilbank D. G., Loubser S. I., 2017, MNRAS, 467, 4101
Grylls P. J., Shankar F., Zanisi L., Bernardi M., 2019, MNRAS, 483, 2506 (Paper I)
Guo Q. et al., 2011, MNRAS, 413, 101
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
Hopkins P. F. et al., 2009, MNRAS, 397, 802
Hopkins P. F., Younger J. D., Hayward C. C., Narayanan D., Hernquist L., 2010a, MNRAS, 402, 1693
Hopkins P. F. et al., 2010b, ApJ, 715, 202
Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, Astron. Lett., 44, 8
Lapi A. et al., 2011, ApJ, 742, 24
Lapi A., Mancuso C., Bressan A., Danese L., 2017, ApJ, 847, 13
Leja J., van Dokkum P. G., Franx M., Whitaker K. E., 2015, ApJ, 798, 115
Leja J. et al., 2019, ApJ, 877, 140
Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119
Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2007, MNRAS, 382, 1394
Matharu J. et al., 2019, MNRAS, 484, 595
McAlpine S. et al., 2015, MNRAS, 450, 1937
Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943
Menci N., Gatti M., Fiore F., Lamastra A., 2014, A&A, 569, A37
Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3
Moster B. P., Somerville R. S., Maulbetsch C., Van Den Bosch F. C., Maccì O A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822
Nelson D. et al., 2018, MNRAS, 475, 624
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Planck Collaboration XIII, 2015, A&A, 594, A13
Press W. H., Schechter P., 1974, ApJ, 187, 425
Salpeter E. E., 1955, ApJ, 121, 161
Schawinski K. et al., 2014, MNRAS, 440, 889
Shankar F., 2013, Class. Quantum Gravity, 30, 244001
Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14
Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2011, MNRAS, 428, 109
Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, MNRAS, 128, 109
Shankar F. et al., 2014, MNRAS, 439, 3189
Shankar F. et al., 2015, ApJ, 802, 73
Springel V. et al., 2018, MNRAS, 475, 676
Tomczak A. R. et al., 2014, ApJ, 783, 85
van Dokkum P. G. et al., 2010, ApJ, 709, 1018
van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Wang T. et al., 2016, ApJ, 828, 56
Wen Z. L., Han J. L., 2018, MNRAS, 481, 4158
Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336
Wetzel A. R., Deason A. J., Garrison-Kimmel S., 2015, ApJ, 807, 49
Wright E. L. et al., 2010, AJ, 140, 1868
Yang X., Mo H. J., Van Den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41
Zavala J., Avila-Reese V., Firmani C., Boylan-Kolchin M., 2012, MNRAS, 427, 1503

APPENDIX A: ABUNDANCE MATCHING MCMC

Fig. A1 shows the redshift $z = 0.1$ output from the MCMC abundance matching fits. It becomes immediately obvious that the low-mass slope (β) is poorly constrained however the impact on the SMF is limited within the margin of error. The position of the knee (M) is well constrained against both the normalization (N) and the high-mass slope (γ). The shape of the constraint between the normalization and gamma emanates from the need to produce high-mass galaxies, if the normalization is decreased the slope must increase to ensure enough haloes produce massive galaxies.

Fig. A2 shows the redshift $z > 0.1$ output from the MCMC abundance matching fits. All parameters have low evolution and the SMHM relation evolves only weakly with redshift. For M, β, and γ where the distributions are wide or close to the prior we have tested wider priors and insignificant change is found.

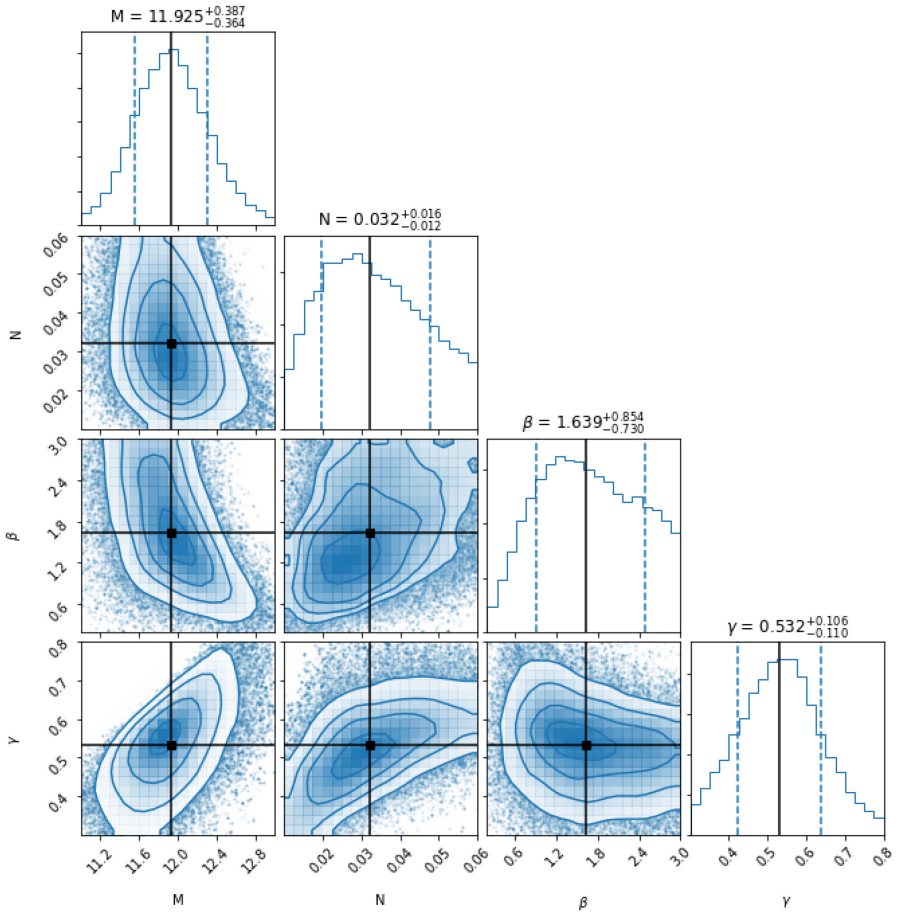


Figure A1. We show the MCMC parameter space for the redshift $z = 0.1$ fit. The position of the knee (M), the normalization (N), the low-mass slope (β), and the high-mass slope (γ) are shown from left to right. Columns are titled with the best-fitting values and 16th/84th percentile errors. The black lines show the best-fitting value with a black square at intersections, the 16th/84th percentiles are shown with blue dashed lines on the histograms.

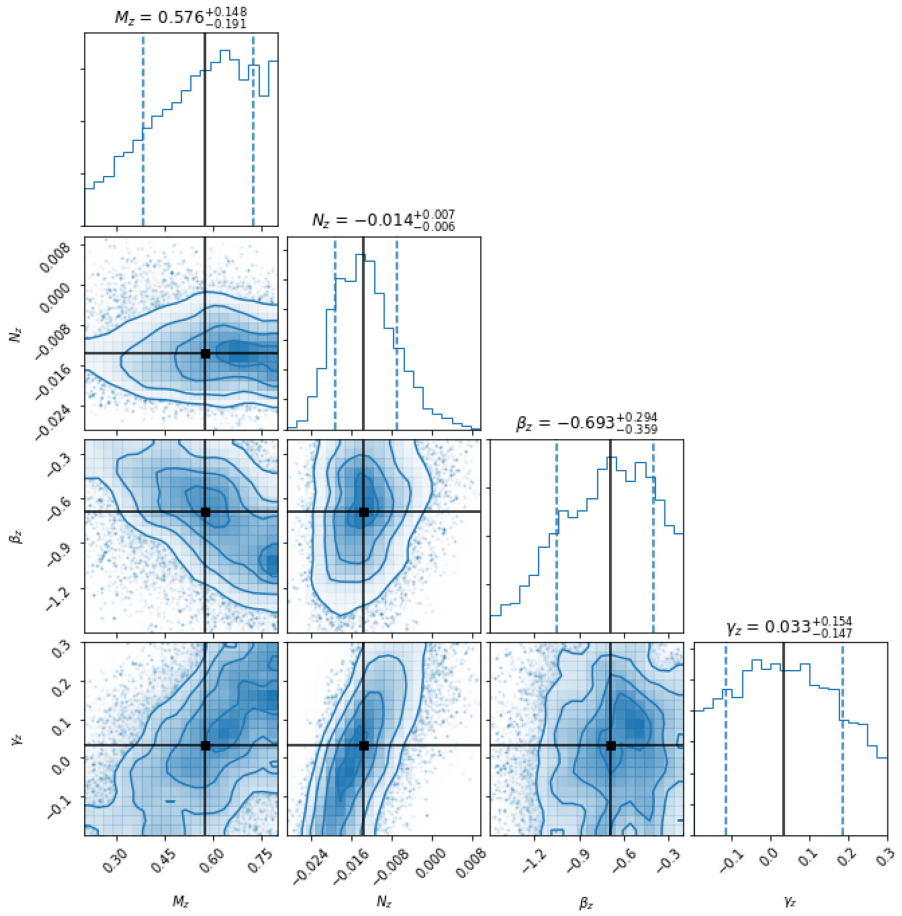


Figure A2. We show the MCMC parameter space for the high redshift $z > 0.1$ fit. The evolution of: the position of the knee (M_z), the normalization (N_z), the low-mass slope (β_z), and the high-mass slope (γ_z) are shown from left to right. Columns are titled with the best-fitting values and 16th/84th percentile errors. The black lines show the best-fitting value with a black square at intersections, the 16th/84th percentiles are shown with blue dashed lines on the histograms.

APPENDIX B: FITTING THE STAR FORMATION RATES

The fit to the SFR we derive in Section 4.3 is given by the following, equation B1,

$$\mathrm{SFR}(M_*, z) = 2N(z)\left[\left(\frac{M_*}{M_\mathrm{n}(z)}\right)^{-\alpha(z)} + \left(\frac{M_*}{M_\mathrm{n}(z)}\right)^{\beta(z)}\right]^{-1}$$

$$\log_{10} N(z) = 10.65 + 0.33z - 0.08z^2$$

$$\log_{10} M_\mathrm{n}(z) = 0.69 + 0.71 * z - 0.088z^2$$

$$\alpha(z) = 1.0 - 0.022z + 0.009z^2$$

$$\beta(z) = 1.8 - 1.0 * z - 0.1z^2. \quad \text{(B1)}$$

APPENDIX C: COMPARISON OF IN SITU/EX SITU GROWTH TO OTHER MODELS

In Fig. C1 we show the *in situ* versus *ex situ* growth with the same model as shown in Fig. 6, we add to this plot data extracted from the Illustris TNG100 simulation. In Fig. 3 we see Illustris has a shallower slow mass slope and a steeper high mass slope such that more stellar mass is mapped into haloes of all sizes. We see the change in both of these slopes reflected in the accretion histories, first, for the lower mass galaxies (see $\log 10 M_{*,\mathrm{cen}} = 10^{11}$) closer to the SMHM knee we find enhanced accretion due to the larger masses from more minor mergers. Secondly the high-mass slope is a direct result of the accretion, to support the same merger assembly with the higher mass galaxies in the satellite haloes above the knee where galaxy growth is dominated by the accretion the galaxy growth with halo size must be enhanced.

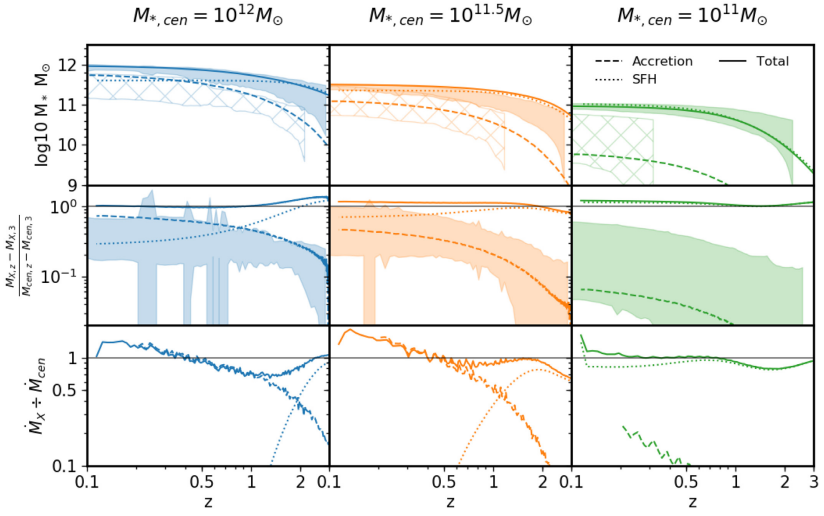


Figure C1. As in Fig. 6 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,\mathrm{cen}} = 10^{12}$, $10^{11.5}$, and 10^{11} ($\mathrm{M}_\odot$) in blue, orange, and green, respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies, the SFR, and the mass growth rate of the central galaxy predicted by abundance matching. In the top panel the shaded regions are galaxies selected from the Illustris simulation; the hashed region is then the satellite accretion from Illustris, in the middle panel the shaded region is the ratio of satellite accretion from Illustris. The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should therefore be close to or on these lines.

In Fig. C2 we show the *in situ* versus *ex situ* growth with the same model as shown in Fig. 6, we add to this plot data from the EMERGE model from Moster et al. (2018) shown as black lines. The solid lines show the total galaxy mass followed back selecting populations by mass at redshift $z = 0.1$. The dotted and dashed lines show the amount of galaxy mass formed *in situ* and *ex situ*, respectively. In all cases EMERGE predicts satellite accretion becomes the dominant mass growth pathway at higher redshifts than STEEL. In the third column we see that EMERGE and STEEL also disagree about the mass growth history of $\log10M_{*,\rm cen} = 11$ galaxies, however, both models agree that the dominant mass growth path of galaxies at this mass are *in situ* processes.

In Fig. C3 we show for the $\log10M_{*,\rm cen} = 11.5$, and 11 galaxies the central growth and SFR ratio from Behroozi et al. (2019). The central growth is close to that found from STEEL and the SFR transition for $\log10M_{*,\rm cen} = 11.5$ is an excellent match.

In Fig. C4 we show a comparison with the semi-analytic model described in Menci et al. (2014). At all masses the stellar growth is substantially different to STEEL and the other models shown in this appendix. Furthermore the semi-analytic model shows little change in the accreted mass ratio over cosmic time, again this is inconsistent with the findings from STEEL and the other models presented in this section. We attribute most of the differences seen here to the substantial difference in the SMHM relationship shown in Fig. C5.

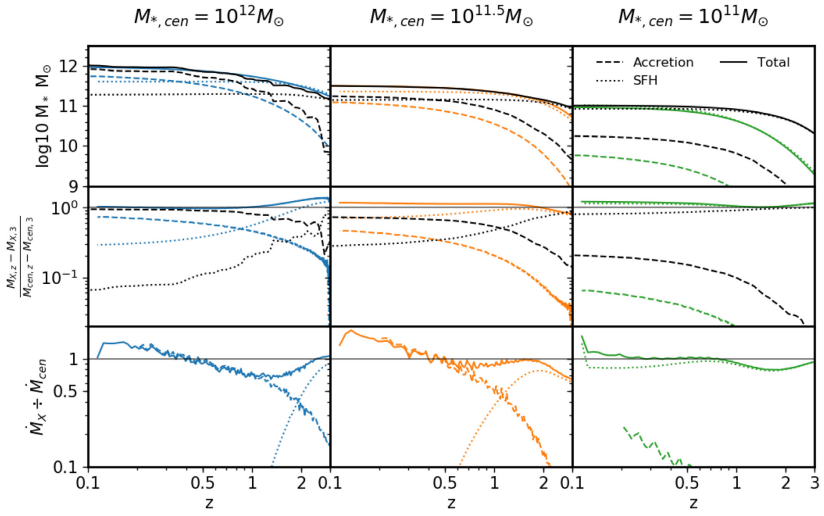


Figure C2. As in Fig. 6 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,\rm cen} = 10^{12}$, $10^{11.5}$, and 10^{11} ($M_\odot$) in blue, orange, and green, respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies, the SFR, and the mass growth rate of the central galaxy predicted by abundance matching. In the top and middle rows we add black lines to show the *in situ* and *ex situ* growth from EMERGE Moster et al. (2018). The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should therefore be close to or on these lines.

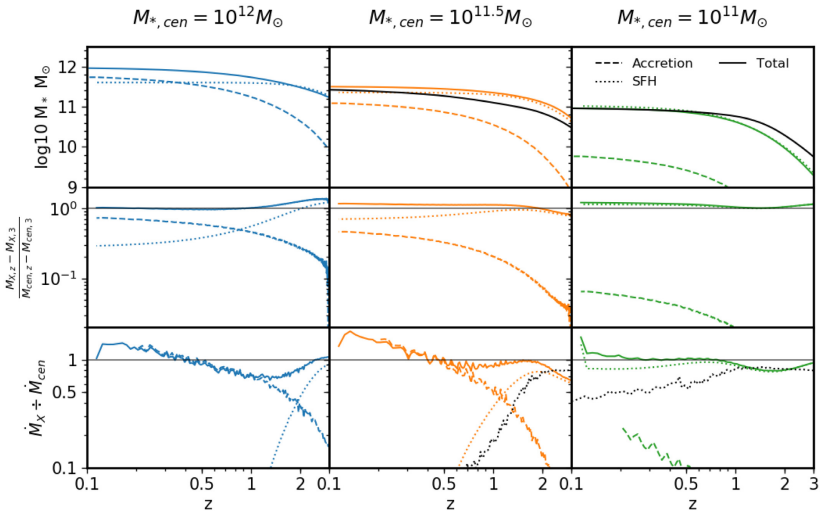


Figure C3. As in Fig. 6 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,\mathrm{cen}} = 10^{12}$, $10^{11.5}$, and 10^{11} ($\mathrm{M}_\odot$) in blue, orange, and green, respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies, the SFR, and the mass growth rate of the central galaxy predicted by abundance matching. In the top and bottom rows, for the $\log 10 M_{*,\mathrm{cen}} = 11.5$, and 11, we add black lines to show the central galaxy growth and the SFR ratio from Behroozi et al. (2019). The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should therefore be close to or on these lines.

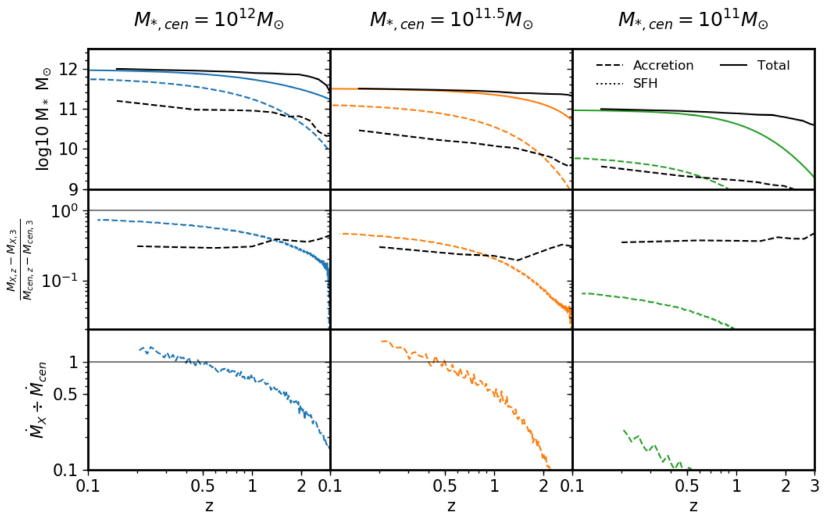


Figure C4. As in Fig. 6 three 'mass tracks' are shown that have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} ($M_\odot$) in blue, orange, and green, respectively. The satellite galaxy accretion is shown for evolved satellites with a dashed line and the mass from star formation shown with a dotted line. The top panel shows the total mass of the central (solid line) and the total mass gained from accretion or star formation. The middle panel shows the fraction of the total galaxy mass formed from satellite accretion or star formation since redshift $z = 3$. The bottom panel shows the ratio of the mass accretion rate from satellite galaxies, the SFR, and the mass growth rate of the central galaxy predicted by abundance matching. In the top and middle rows we add black lines to show the *ex situ* growth and central growth from the semi-analytic model described in Menci et al. (2014). The grey lines in the second and third panel are at unity, the solid lines showing the sum of the other two factors should therefore be close to or on these lines.

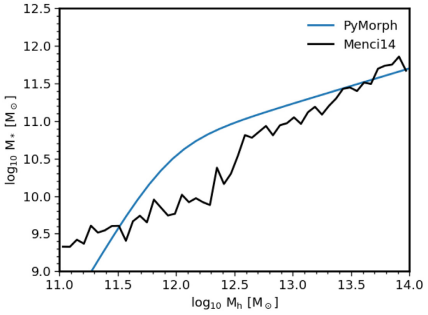


Figure C5. We show the SMHM relationship for PYMORPH (blue line, used in this work) and for the semi-analytic model from Menci et al. (2014) (black line).

APPENDIX D: HALO MERGER RATE IN THE MILLENNIUM COSMOLOGY.

The statistical dark matter accretion history used in STEEL is cosmologically flexible.[12] The current prescriptions for the halo growth histories and dark matter substructure (van den Bosch et al. 2014; Jiang & van den Bosch 2016) can be used for ΛCDM models that are within a factor two of current constraints. Using COLLOSUS (Diemer 2017) we set the cosmology used in STEEL to that of the Millennium simulation ($\Omega_m = 0.25$, $\Omega_b = 0.045$, $\Omega_\Lambda = 0.75$, $h = 0.73$). The statistical dark matter accretion history is then recalculated. Fig. D1 shows the halo merger rate from STEEL from this alternative accretion history. The halo merger rate tracks from Fakhouri & Ma (2010) shown in Figs 4 and D1 are calculated using the Millennium simulation cosmology. The deviation between the merger rate of STEEL and that of Fakhouri & Ma (2010) still remains despite the change of cosmology, as discussed in Section 3.1 this is an effect of the algorithms used to build halo merger trees.

[12]The eventual goal of STEEL is to be cosmologically independent such that any cosmology can be used.

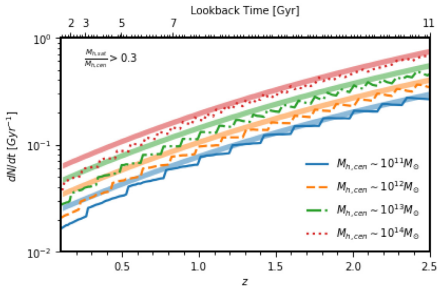


Figure D1. The evolution of merger rate per Gyr at fixed halo mass. Lines are from STEEL with cosmology altered to that of the analytic fits from Fakhouri et al. (2010) (shaded bands). Halo masses shown are $M_{\rm h,cen}$: 10^{11}, 10^{12}, 10^{13}, $10^{14}\,{\rm M}_\odot\,h^{-1}$ as labelled.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **483,** 2506–2523 (2019) doi:10.1093/mnras/sty3281
Advance Access publication 2018 December 3

A statistical semi-empirical model: satellite galaxies in groups and clusters

Philip J. Grylls [1]★ F. Shankar,[1] L. Zanisi[1] and M. Bernardi[2]

[1]*Department of Physics and Astronomy, University of Southampton, Highfield SO171BJ, UK*
[2]*Department of Physics and Astronomy, University of Pennsylvania, PA 19104, USA*

Accepted 2018 November 28. Received 2018 November 27; in original form 2018 July 27

ABSTRACT

We present STEEL, a STatistical sEmi-Empirical modeL, designed to probe the distribution of satellite galaxies in groups and clusters. Our fast statistical methodology relies on tracing the abundances of central and satellite haloes via their mass functions at all cosmic epochs with virtually no limitation on cosmic volume and mass resolution. From mean halo accretion histories and subhalo mass functions, the satellite mass function is progressively built in time via abundance matching techniques constrained by number densities of centrals in the local Universe. By enforcing dynamical merging time-scales as predicted by high-resolution *N*-body simulations, we obtain satellite distributions as a function of stellar mass and halo mass consistent with current data. We show that stellar stripping, star formation, and quenching play all a secondary role in setting the number densities of massive satellites above $M_* \gtrsim 3 \times 10^{10}\,\mathrm{M}_\odot$. We further show that observed star formation rates used in our empirical model over predict low-mass satellites below $M_* \lesssim 3 \times 10^{10}\,\mathrm{M}_\odot$, whereas star formation rates derived from a continuity equation approach yield the correct abundances similar to previous results for centrals.

Key words: galaxies: clusters: general – galaxies: formation – galaxies: groups: general – galaxies: haloes.

1 INTRODUCTION

Cold dark matter (CDM) predicts a hierarchical assembly of dark matter haloes, which grow in mass through accretion from smaller units to increasingly larger systems. The complex mass assembly of haloes is recorded in the dark matter merger trees. After infall into larger systems, accreted satellite haloes will eventually merge with the parent halo through stripping and orbital decay from dynamical friction. Dark matter structure formation has been studied in numerous numerical works with (e.g. Lemson & Consortium 2006; Klypin et al. 2016) and without (e.g. Vogelsberger et al. 2014; Fattahi et al. 2016) the baryonic component. Dark matter-only simulations have been performed on very large scales, with cosmological volumes up to several Gpc^3 (Potter, Stadel & Teyssier 2017). In comparison, due to the computational complexities introduced by hydrodynamics, simulations with baryons have been limited to orders of magnitude smaller volumes, around a hundred Mpc^3. The discrepancy in simulated volumes inevitably leads to structures with intrinsically lower number densities limiting their usefulness in studying less common but important galaxy populations, such as massive satellite galaxies residing in relatively small parent haloes. Hydrodynamical simulations have resolution and time constraints imposed by the computational power they can leverage, ultimately being forced to turn to sub-grid analytic physical recipes below the particle or mesh resolution.

Semi-analytic models (e.g. Granato et al. 2004; Baugh 2006; Monaco, Fontanot & Taffoni 2007; Hirschmann et al. 2012; Shankar et al. 2013) make comprehensive predictions on the evolutionary patterns of galaxies by following the growth and assembly along dark matter merger trees. Semi-analytic models are necessarily characterized by a number of input assumptions and free parameters associated with different formation mechanisms. Semi-analytic models make the admirable effort towards a holistic view of galaxy formation, but they are known to suffer from degeneracies in assumptions and related parameters (e.g. González et al. 2011; Lapi & Cavaliere 2011), and could still be limited by volume effects.

In the past decades, much attention has been devoted to probing the connection between dark matter and galaxies in a more empirical fashion, attempting to build a self-consistent picture of dark matter/galaxy assembly and distribution making use of more observationally driven methods. Halo occupation distribution models were introduced as a first step in this direction (e.g. Berlind & Weinberg 2002; Cooray & Sheth 2002). These models are designed to predict the average number of central and satellite galaxies in haloes by simultaneously matching observed number densities and clustering at a given epoch. Halo occupation models by construction do not account for the full assembly history of the galaxies and may be susceptible to assembly bias effects (e.g. Dalal et al. 2008; Zentner, Hearin & van den Bosch 2014).

★ E-mail: p.grylls@soton.ac.uk

Abundance matching between the galaxy luminosity/stellar mass function (SMF) and the halo mass function (HMF) is a semi-empirical technique that has been widely adopted in the last two decades to constrain, on a statistical basis, the mean relation between stellar mass and host halo[1] mass (e.g. Kravtsov et al. 2004; Vale & Ostriker 2004; Yang et al. 2004; Shankar et al. 2006). In its simplest form, abundance matching relies on the basic idea that larger galaxies reside in larger dark matter haloes, and relative abundances are used to create a (mean) stellar mass–halo mass (SMHM) mapping $M_* = \mathrm{SMHM}(M_\mathrm{h}, z)$ at any relevant epoch (Yang, Mo & Van den Bosch 2003; Behroozi, Conroy & Wechsler 2010; Moster et al. 2010). This technique has been used as a flexible tool to predict the full mass and structural assembly history of galaxies along dark matter merger trees in a full cosmological context.

Semi-empirical models (e.g. Hopkins et al. 2009b; Cattaneo et al. 2011; Zavala et al. 2012; Shankar et al. 2014a) were conceived as models where galaxies are assigned to haloes at any epoch along dark matter merger trees via abundance matching relations, thus allowing to track the full stellar mass evolution in, for example, the main progenitor branch of the merger tree. Semi-empirical models, although unavoidably more restrained in scope, avoid the often heavy parametrizations incorporated in semi-analytic models, as galaxy stellar masses and other properties (e.g. disc sizes) are empirically assigned to dark matter haloes. In a semi-empirical approach, a minimum number of free parameters are included in the models allowing exploration of specific aspects of galaxy formation minimizing the danger of degeneracies. Despite their flexibility, traditional semi-empirical models can be subject to volume limitation effects due to the finite number of merger trees and/or dark matter haloes. Both semi-analytic and semi-empirical models can therefore suffer from low number statistics especially at high stellar masses when built on top of dark matter simulations.

In this work, we present STEEL, a STastical sEmi-Empirical modeL, built using a novel, fast and flexible methodology to probe the number densities and average properties of galaxies at any given stellar mass, epoch and environment, with negligible computational limitation on volume or mass resolution. This is particularly useful when aiming at simultaneously probing, as in this work, the distribution of massive centrals as well as massive satellites in lower mass haloes. The method employed in this paper starts from tracking backwards in time the mean accretion histories of small bins in halo mass, weighted by the HMF. At each time-step, satellite galaxy number densities are associated with the main halo mass bin via the corresponding time growth in subhalo mass function (SHMF). The main aim of this work is to predict satellite number densities as a function of stellar mass and halo mass. We will show that the dominant parameter controlling the abundance of massive satellites is the dynamical friction time-scale. Nevertheless, the evolution of satellite galaxies, as central ones, is of course a result of a combination of star formation, quenching, and stellar stripping. We will probe the (limited) impact on our results of each one of these processes.

In what follows, we adopt the Planck cosmology with $(\Omega_\mathrm{m}, \Omega_\Lambda, \Omega_\mathrm{b}, h, n, \sigma_8) = (0.31, 0.69, 0.05, 0.68, 0.97, 0.82)$.[2] We use M_h and M_* to differentiate between halo and stellar mass, respectively; we name $M_\mathrm{h,cent}$ and $M_\mathrm{h,sat}$ central and satellite, respectively (we direct readers to Appendix C for a complete glossary of the acronyms used throughout this work). Halo masses are defined as virial masses, unless stated otherwise. Wherever relevant, we always assume a Chabrier (2003) stellar initial mass function.

2 DATA

In this work, we take as a reference the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7) from Meert, Vikram & Bernardi (2015, 2016), with improved galaxy photometry. In brief, the data are originally from the SDSS-DR7 spectroscopic sample (Abazajian et al. 2009) containing ∼670 000 galaxies fitted with a Sérsic+exponential model (PyMorph; Meert et al. 2015, and references therein), with associated halo masses (Yang et al. 2012) updated to the new photometry. The improved photometry provides a number of advantages. For instance, the two-dimensional fit more accurately captures the high-mass end of the SMF (Bernardi et al. 2013, 2017a,b) and uses the mass-to-light ratio from Mendel et al. (2014).

At higher redshifts ($0.3 < z < 3.3$), SMFs are from the COSMOS2015 catalogue (Davidzon et al. 2017). In this survey, masses are derived from multiwavelength spectral energy distribution fitting, including ultra-deep and infrared photometry. We note Davidzon et al. (2017) use Bruzual & Charlot (2003) stellar population synthesis models that are consistent within $\lesssim$0.15 dex with the mass-to-light ratios adopted in the SDSS catalogue. Both data sets assume a Chabrier (2003) stellar initial mass function.

3 METHODS: A NEW 'STATISTICAL' APPROACH TO SEMI-EMPIRICAL MODELLING

Semi-analytic models continue to provide invaluable constraints on the formation and evolution of galaxies. They admirably attempt to analytically model from zero order the full evolution of galaxies since the Big Bang to the local Universe (e.g. Menci et al. 2006). This inevitably requires specific input assumptions and related parameters on, for example, the rate of star formation at different epochs and/or environments.

Semi-empirical models are instead characterized by a 'bottom-up' approach, where the least possible assumptions and associated parameters are initially included in the models. Gradually, additional degrees of complexities can be included in the model, wherever needed. Semi-empirical models in this respect are a very powerful complementary tool to semi-analytic models or even cosmological hydrodynamic simulations. Their goal is in fact to provide unique constraints to only specific aspects of galaxy evolution, such as predicting, as in this work, the distribution of satellites in a given dark matter parent halo.[3]

Semi-empirical models (Hopkins et al. 2009b; Cattaneo et al. 2011; Zavala et al. 2012; Shankar et al. 2014a) are broadly characterized by the following steps:

[1]Here and throughout this work, 'host halo' refers to the dark matter (sub)halo associated with a given galaxy residing at the (sub)halo's centre.

[2]We note that Planck's best-fitting cosmology has slightly different parameters than those adopted in some of the observational probes included in this work, such as the SMFs ($\Omega_\mathrm{m}, h = (0.30, 0.70)$]. However, we have checked that correcting the SMF volumes and luminosities to the same Planck's cosmology yields essentially indistinguishable results in the implied SMHM relation. We will thus ignore such tiny differences in cosmological parameters in what follows. We choose to retain the halo cosmology as the full dark matter backbone of the model presented in this work is built and calibrated against the Planck Cosmology (Planck Collaboration et al. 2016).

[3]Here and throughout this work, 'parent halo' refers to the central halo containing subhaloes/satellite galaxies.

(i) Central galaxies, associated with the main progenitor branch[4] of the host dark merger tree, are initially defined as gas-rich with disc like profiles at early epochs, as supported by deep *HST* observations in CANDELS (e.g. Huertas-Company et al. (2015, 2016)). All their stellar mass and structural properties are assigned via empirical, time-dependent relations, more significantly the SMHM relation (e.g. Moster, Naab & White 2013).

(ii) Central galaxies are then re-initialized at each time-step during the evolution and can gradually transform their morphology via 'in situ' processes, such as (more or less) violent disc instabilities, and/or 'ex situ' processes, such as mergers.

(iii) Satellite galaxies are those associated with each dark matter branch merging to the main progenitor. They are assigned all the mass and structural properties of a central galaxy in a typical central halo of the same mass at the time of infall. Satellites may further evolve in stellar mass, given their specific star formation rate (SFR) at infall, possibly accompanied by some stellar stripping.

The typical semi-empirical model described above is a powerful and very competitive tool to explore trends in the expected mass and/or structural evolution of galaxies in a given interval of stellar mass for a given set of parameters in, e.g., amount of stripping and/or orbital energy (Shankar et al. 2015). Based on, by design, closely following the evolution of each dark matter merger tree, the basic semi-empirical approach discussed above is still limited by the resolution and volume of the background dark matter simulation. This could be particularly restraining and computationally expensive especially when modelling very massive galaxies and their accompanying large volumes.

In this paper, we propose a statistically based semi-empirical approach. We explore a large range of stellar masses, with virtually no limitation in sample size, and still retain the high degree of flexibility characteristic of a semi-empirical model. The core steps in building the dark matter backbone of STEEL are summarized as follows (see diagram in Fig. 1):

(i) At the redshift of interest $\bar{z}$, we start from the HMF to compute the abundances of (parent/central) haloes (panel A).

(ii) Each parent/central halo mass is then followed backwards in time following its average mass growth history, $< M_{\rm halo}(\bar{z}) >$, as expected from extended Press–Schechter (Press & Schechter 1974) and/or numerical simulations (panel B).

(iii) At each time-step, we then calculate the difference in subhalo population between z and $z + {\rm d}z$ to estimate the expected average number density and masses of subhaloes accreted on to the main progenitor in the redshift interval ${\rm d}z$ (panel D).

We will show how using analytic HMFs and a 'statistical accretion history', in place of halo catalogues and merger trees, we are able to simultaneously examine massive central haloes, and massive satellites in lower mass parents. The latter are usually characterized by low number densities and can be omitted in more traditional approaches that rely on cosmological volumes. For example, massive satellite galaxies with stellar mass $M_* > 10^{11}{\rm M}_\odot$ residing in relatively smaller parent haloes of mass $M_{\rm h} \sim 10^{12.75}\,{\rm M}_\odot$ have predicted number densities of $\sim 10^{-6}\,{\rm Mpc}^{-3}$ (see Fig. 11, panel 3), which would correspond to $\sim$1 galaxy in a volume of $\sim$100 $({\rm Mpc/h})^3$, a typical volume adopted in state-of-the-art cosmological hydrodynamical simulations. On the other hand, even in larger dark matter-only simulations such as the Bolshoi [250 $({\rm Mpc/h})^3$] (Klypin et al. 2016), one would expect at the most $\sim$16 of these galaxies. As massive satellite are so rare in other simulations, a robust statistical determination of abundances is therefore vital to fully constrain the connection between galaxy formation and hierarchal assembly predicted by ΛCDM cosmology at all mass ranges (Neistein et al. 2013).

In this section, we detail each step of this statistically based semi-empirical approach. We start by outlining in Section 3.1 the analytic model of dark matter accretion histories characterizing the mean growth of the main progenitor halo and accreted subhaloes including subhalo merging times. We then discuss in Section 3.2 how we populate dark matter haloes with galaxies, using the SMHM relation for central galaxies and its evolution with redshift. Finally, in Section 3.3, we discuss the evolutionary processes that affect the satellite galaxies whilst they reside in their dark matter parent haloes. The parameters used in this section and their origins are given in Table 1.

3.1 Statistical dark matter accretion history

The core principle of our methodology is to treat parent haloes, and satellites galaxies/haloes, as 'average' populations avoiding issues with volume and resolution as described previously. Specifically, the goal of this paper is to reproduce the observed distribution of $z = 0.1$ satellite galaxies with mass $M_* > 10^9{\rm M}_\odot$. In this subsection, we detail step-by-step construction of the backbone of STEEL complemented with a graphic representation in Fig. 1.

3.1.1 Central haloes

We start by considering a fine grid of central dark matter haloes ranging from $M_{\rm h} = 10^{11}\,{\rm M}_\odot$ to $M_{\rm h} = 10^{15}\,{\rm M}_\odot$ at redshift $z = 0$. Their number densities are given by the HMF, which is obtained using HMFcalc (Murray, Power & Robotham 2013). HMFcalc includes a number of different mass functions. For this work, we use the HMF from Tinker et al. (2010).[5] The HMF provides the number densities of haloes in a given mass bin (Fig. 1, panel A). The average mass growth histories of all main progenitors with mass in the bin of halo mass $(M_{\rm h}, M_{\rm h} + {\rm d}M_{\rm h})$ are then calculated using the analytic model from Bosch et al. (2014).[6] This provides the average 'main progenitor' branch of a traditional merger tree for each mass bin at $z = 0$ (Fig. 1, panel B).

3.1.2 Assigning subhaloes to parent haloes

In order to predict the number of satellite galaxies, we must associate to each parent/central halo the number and mass of subhaloes they are expected to contain. To achieve this, we use the SHMF. The SHMF describes the expected distribution of subhaloes of mass $M_{\rm h,sat}$, in a given parent halo of mass $M_{\rm h,cent}$, as a function of $M_{\rm h,sat}/M_{\rm h,cent}$. Multiple definitions for the SHMF exist depending

[4]The main progenitor is defined in traditional merger trees as the largest progenitor halo from any given halo merger, usually the largest halo at any time-step. In this work shown in panel B of Fig. 1, we simply have the average growth of this halo but it is convenient to visualize in the traditional sense.

[5]We adopt the COLOSSUS PYTHON package (Diemer 2017) for all halo mass conversions required.

[6]This model further improves on the seminal work by Parkinson, Cole & Helly (2008), which was aimed at reproducing numerical merger trees, optimized with small redshift steps minimizing the development of systematic errors at late cosmic epochs.

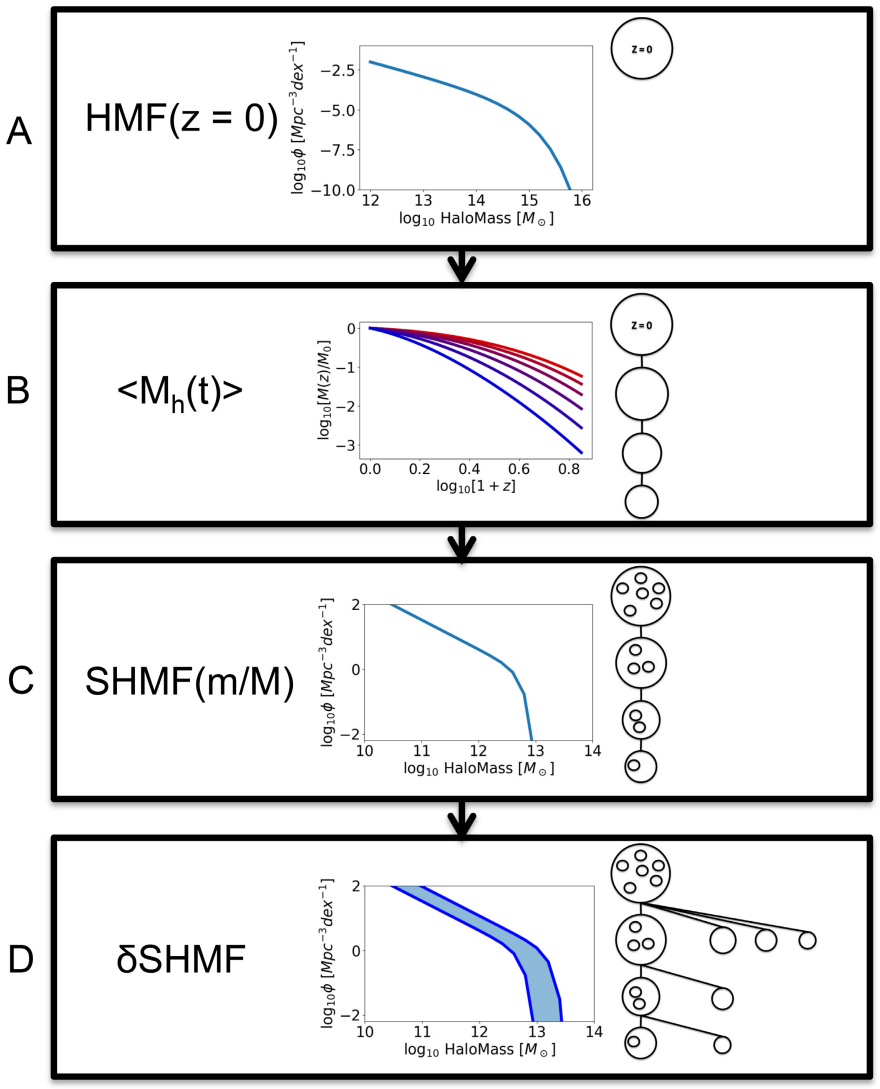


Figure 1. We show the main steps in building the statistical dark matter backbone of STEEL described in Section 3.1. Each panel shows the feature from a traditional merger tree and the statistical function used to replace it. (A) The HMF is used to calculate the number densities of central haloes. (B) Average mass growth histories calculate the size of each mass bin at previous epochs. (C) The SHMF is used to populate each central at each redshift with subhaloes. (D) The average number densities of accreted subhaloes at each epoch are calculated by comparing the SHMF over consecutive redshift steps.

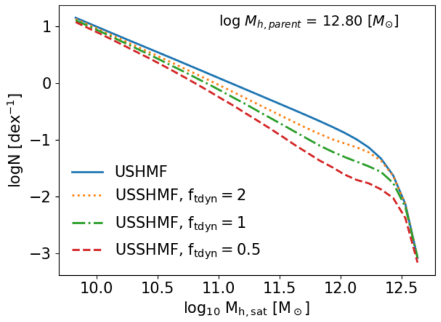


Figure 2. Comparison between the USHMF (solid line) and three USSHMF (dotted lines) for a parent halo of mass log $M_{\rm h,parent}$ [M$_\odot$] = 12.80. The factor $f_{t_{\rm dyn}}$ is applied to the merging time-scales of the haloes. Lower factors correspond to lower USSHMF, where more sub-haloes have merged.

on the way a subhalo is defined. In this work, we use two definitions of the SHMF. The first is the unevolved SHMF (USHMF), which describes the total subhaloes accreted over a parent halo's lifetime. In the USHMF, any merging or stripping in the subhaloes occurring after infall are ignored. A number of groups have been able to constrain the USHMF (Giocoli, Pieri & Tormen 2008; Jiang & van den Bosch 2016). In what follows, we use the latest rendition of the USHMF by Jiang & van den Bosch (2016), which is calibrated against the Bolshoi simulation.[7]

The second definition we use in this work is the unevolved 'surviving' SHMF (USSHMF). Subhalo masses are assumed frozen at infall, but the subhalo number densities can reduce compared to the USHMF as the USSHMF accounts for subhalo disappearance due to tidal disruption in the parent halo. We show in Fig. 2 for a representative parent halo of mass log $M_{\rm h,cent}$M$_\odot$ = 12.80, the USHMF and three USSHMF characterized by different dynamical friction time-scales $\tau_{\rm dyn}$, as described in Section 3.1.4. Larger $\tau_{\rm dyn}$ lead to a milder reduction in subhalo number densities as subhaloes take longer to merge with the parent halo. Lower $\tau_{\rm dyn}$ are less effective in reducing the number densities of smaller subhaloes that are more likely to have dynamical friction time-scales comparable to or larger than the Hubble time at $z = 0$.

3.1.3 Average subhalo accretion

At each redshift step along the mass growth histories, we calculate the USHMF associated with the parent halo mass. What we create is equivalent to the substructure found in traditional merger trees (Fig. 1, panel C). However, unlike traditional methods, our statistical approach is able to probe 'rare' subhaloes without running prohibitively large volumes of merger trees.

For each time-step, we can now calculate a mass function describing the number density of subhaloes accreted on to the population of central haloes in the halo mass bin [$M_{\rm h,cent}(z)$, $M_{\rm h,cent}(z)$ + $dM_{\rm h,cent}(z)$]. The latter is achieved by differentiating the USHMF across two neighbouring redshift steps z and $z + {\rm d}z$, we can calculate the average number density of subhaloes of any given mass $M_{\rm h,sat}$ that are accreted in the redshift interval dz on to the main progenitor haloes with mass in the bin [$M_{\rm h,cent}(z)$, $M_{\rm h,cent}(z) + dM_{\rm h,cent}(z)$],

$$\delta{\rm USHMF}[z,\, M_{\rm h,cent},\, M_{\rm h,sat}] \\ . \;= {\rm USHMF}\Big(\frac{M_{\rm h,sat}}{M_{\rm h,cent}(z)}\Big) - {\rm USHMF}\Big(\frac{M_{\rm h,sat}}{M_{\rm h,cent}(z+\delta z)}\Big) \qquad (1)$$

In this way, the unevolved subhalo accretion history (δUSHMF) is retrieved for all main progenitor haloes at all redshifts.

3.1.4 Subhalo merging time-scale

From the unevolved subhalo accretion history, we need to isolate which subhalo mass bins survive at each epoch. The sum of all the surviving subhaloes (at each epoch) then yields the USSHMF. A key parameter used to calculate the USSHMF is the 'observability time-scale' (or survival time) of each subhalo mass bin [$M_{\rm h,sat}(z)$, $M_{\rm h,sat}(z)$ + $dM_{\rm h,cent}(z)$] associated with a parent halo mass bin [$M_{\rm h,cent}(z)$, $M_{\rm h,cent}(z) + dM_{\rm h,cent}(z)$]. This time-scale is equivalent to the merger time-scale $\tau_{\rm merge}$ of a subhalo of mass $M_{\rm h,sat}$ in a parent halo mass $M_{\rm h,cent}$. To calculate $\tau_{\rm merge}$, we use the routines in equation (2), derived from N-body simulations (Boylan-Kolchin, Ma & Quataert 2008)

$$\tau_{\rm merge} = (f_{t_{\rm dyn}}\tau_{\rm dyn})\frac{A(M_{\rm h,cent}/M_{\rm h,sat})^B}{\ln(1+M_{\rm h,cent}/M_{\rm h,sat})}\exp\Big(C\frac{J}{J_c(E)}\Big)\Big(\frac{r_c(E)}{r_{\rm vir}}\Big)^D, \qquad (2)$$

where $A = 0.9$, $B = 1.0$, $C = 0.6$, $D = 0.1$ (McCavana et al. 2012). The factor $\tau_{\rm dyn}$ is given by (Jiang & van den Bosch 2016)

$$\tau_{\rm dyn} = 1.628h^{-1}{\rm Gyr}\Big(\frac{\Delta_{\rm vir}(z)}{178}\Big)^{-\frac{1}{2}}\Big(\frac{H(z)}{H_0}\Big)^{-1}. \qquad (3)$$

Our method of considering average halo mass and accretion histories does not allow tracking single orbits and associated orbital energies. We assume instead an average orbit circularity of 0.5 (Khochfar & Burkert 2006), thus reducing the dependence on the angular momentum and radial components, $\frac{J}{J_c(E)}$ and $\frac{r_c(E)}{r_{\rm vir}}$, to a constant. In other words, this approximation is consistent with the approach of taking the average expected orbits of subhaloes at fixed parent halo mass. The key parameter of our analysis is the factor $f_{t_{\rm dyn}}$ included in equation (2). The fudge factor $f_{t_{\rm dyn}}$ takes into account the systematic uncertainties induced by numerical resolution effects in N-body simulations that are unable to resolve the full merging time-scales of subhaloes and/or the satellite galaxies they host (Bosch et al. 2018). The parameter $f_{t_{\rm dyn}}$ increases or decreases the dynamical times of 'merging' satellites enabling an exploration of the effect of dynamical time on the final number density distributions of satellite galaxies at any given epoch.

3.1.5 Surviving subhalo population

At each redshift, we can now use the unevolved subhalo accretion history and the observability time-scale $\tau_{\rm merge}$ to calculate the total observable subhalo population associated with any given parent halo mass bin [$M_{\rm h,cent}(z)$, $M_{\rm h,cent}(z) + {\rm d}M_{\rm h,cent}(z)$], i.e. the USSHMF (shown by the dashed lines in Fig. 2). To compute the implied total number densities of unmerged subhaloes with mass [$M_{\rm h,sat}(z)$, $M_{\rm h,sat}(z) + {\rm d}M_{\rm h,sat}(z)$] at any redshift of interest, we convolve the

[7] We direct the interested reader to Jiang & van den Bosch (2016) for further discussion of the USHMF as well as of other SHMFs, such as the evolved SHMF where the number densities are effected by both subhalo stripping and mergers.

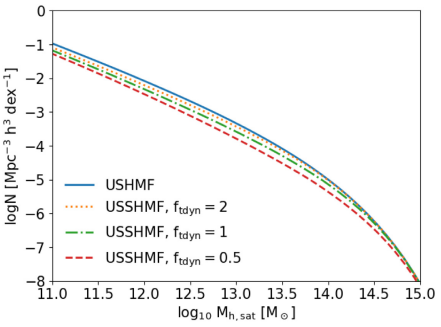


Figure 3. Example of the 'total' USHMF (solid line) and three 'total' USSHMF (dotted lines) corresponding to three different $f_{t_{\rm dyn}}$ factors.

USSHMF with the HMF,

$$N(M_{\rm h,sat},z)=\int \mathrm{USSHMF}\left(\frac{M_{\rm h,sat}}{M_{\rm h,cent}}\right)\mathrm{HMF}(M_{\rm h,cent},z)\mathrm{d}M_{\rm h,cent}. \tag{4}$$

Fig. 3 shows the total observable subhalo population for different $f_{t_{\rm dyn}}$ similar to the SHMF in Fig. 2. Furthermore, via appropriate abundance matching algorithms detailed below, we can assign corresponding satellite galaxies to the unevolved subhalo accretion history and obtain the distribution of satellites in a given parent halo mass bin $[M_{\rm h,cent}(z), M_{\rm h,cent}(z)+\mathrm{d}M_{\rm h,cent}(z)]$ by assuming the satellites follow the same merging time-scales as their host subhaloes. Finally, we convolve the latter distributions of satellite galaxies with the HMF to create the total observed satellite stellar mass function (SSMF). Both the distribution of satellite galaxies and the satellite SMF can then be used to constrain $f_{t_{\rm dyn}}$ as demonstrated in Section 4.

3.2 Abundance matching: Building the local populations of central and satellite galaxies

We now describe how we populate dark matter haloes with galaxies using the abundance matching technique. To compute the distribution of satellites in parent haloes, we populate subhaloes at infall, following the central SMHM relation. For the latter, we adopt a parametric form similar to that proposed by Moster et al. (2010),

$$\begin{aligned}
M_*(M_{\rm h},z)&=2M_{\rm h}N(z)\left[\left(\frac{M_{\rm h}}{M_n(z)}\right)^{-\beta(z)}+\left(\frac{M_{\rm h}}{M_n(z)}\right)^{\gamma(z)}\right]^{-1},\\
N(z)&=N_0.1+N_z\left(\frac{z-0.1}{z+1}\right),\\
M_n(z)&=M_{n,0.1}+M_{n,z}\left(\frac{z-0.1}{z+1}\right),\\
\beta(z)&=\beta_0.1+\beta_z\left(\frac{z-0.1}{z+1}\right),\\
\gamma(z)&=\gamma_0.1+\gamma_z\left(\frac{z-0.1}{z+1}\right).
\end{aligned} \tag{5}$$

3.2.1 *Local Universe $z = 0.1$*

To constrain the $z = 0.1$ parameters $(M_n, N, \beta, \gamma, \sigma)$, we construct a SMF from the HMF and the SMHM relation with initial parameters close to previous results (Moster et al. 2013; Shankar et al. 2014a; Buchan & Shankar 2016). The HMF provides the number density for each bin of parent halo mass in the range $[M_{\rm h,cent}(z), M_{\rm h,cent}(z) + \mathrm{d}M_{\rm h,cent}(z)]$. Using the SMHM relation, each bin of parent halo mass is associated with a Gaussian distribution of stellar masses with width controlled by the scatter parameter (σ). Multiplying by the halo mass number density, we convert this distribution into galaxy number densities, which are then added to the relevant stellar mass bins of the SMF in construction. This operation is repeated over each bin of the HMF. The resulting SMF is then compared to the SDSS central SMF. We vary the parameters on a grid and minimize root mean square between the constructed SMF and the data. In our abundance matching framework, we specifically distinguish between central and SSMFs by taking advantage of the SDSS galaxy-halo catalogue described in Section 2.

For comparison with previous works, we apply the same method as described above to the total galaxy/halo population using the USSHMF ($f_{t_{\rm dyn}} = 1.0$), the HMF, and the total SMF. Our total result uses single-epoch [8] abundance matching: we apply the SMHM relation to the total HMF at $z = 0.1$ with no consideration of the infall time of subhaloes.

Table 2 reports the best-fitting parameters using the method above, for central galaxies and total galaxy population. In the left-hand panel of Fig. 4, we show with a solid blue line the $z = 0.1$ SMF for SDSS central galaxies only, abundance-matched with the central HMF (no subhaloes). For comparison we also include the abundance matching fits by Moster et al. (2013), Shankar et al. (2017b), and Behroozi et al. (2018, using the 'true' stellar masses, in table J1). The *total* SMF is reproduced by Moster et al. (2013) and Shankar et al. (2017b) without distinguishing between centrals and satellites galaxies in the SMF. We find our final results for the high mass slope to be steeper than the Moster et al. (2013) but very similar to the ones by Shankar et al. (2017b), as expected given that the latter have tuned their fitting routines to the same galaxy SMF as the one adopted in this paper. At the low mass end our SMHM relation tends to appear slightly steeper than previous works. At least with respect to Shankar et al. (2017b), this is mainly caused by the fact that our results are based on the updated surviving SHMF by Jiang & van den Bosch (2016) that is somewhat steeper than the subhalo mass correction suggested by Behroozi, Wechsler & Conroy (2013) and adopted by Shankar et al. (2017b). Our low-mass slope is, as expected, fully consistent with Behroozi et al. (2018) as their subhaloes have been extracted from the Bolshoi–Planck simulation (Klypin et al. 2016). The right-hand panel of Fig. 4 shows the total, central, and SSMFs as extracted from our SDSS catalogues (black circles, blue triangles, and red squares, respectively) compared with the central SMF generated using the central HMF and the relations from Moster et al. (2013), Behroozi et al. (2018), and our best-fitting relations. The solid black line shows our 'total' relationship where the full population of central and satellite galaxies are retrieved with single-epoch abundance matching.

From this comparison, we conclude that for the SMF and HMF used in this work single-epoch abundance matching is acceptable,

[8] A multiepoch approach considers the infall time of subhaloes, and routines are used for stripping, star formation, and quenching of satellite galaxies. The full set of parameters are fit simultaneously (e.g. Moster et al. 2013).

Table 1. Parameters used in the model and how they are constrained.

Parameter	Description	Equation
	Free parameters	
$f_{t\mathrm{dyn}}$	A factor we include to test the impact of shortening or lengthening dynamical times of all satellites. This is a free parameter we use to find the best fit to satellite distributions.	3
	Constrained parameters	
A,B,C,D	Dynamical friction fitting parameters given by McCavana et al. (2012)	2
M_n, N, β, γ	Abundance matching parameters, constrained by the central SMF and HMF at $z = 0$	5
$M_{n,z}$, N_z, β_z, γ_z	Abundance matching evolution parameters, constrained using SMF and the HMF at $z > 0$.	5
s, M, α	Star formation parameters, constrained by observation given in Tomczak et al. (2016). We also give values to fit our continuity derived SFR in Section 3.3.1	6
τ_q, τ_f, M_{cutoff}	Parameters describing the quenching time of satellites. These parameters are all adapted from models in the literature (Wetzel et al. 2013; Fillingham et al. 2016)	8–10
η_{strip}	Stripping fraction from Cattaneo et al. (2011)	15

Table 2. Parameters for equation (5). The $z = 0.1$ parameters are constrained to 2 d.p. or 2 s.f., whichever is smaller. The $z > 0.1$ parameters are constrained to 1 d.p. or 1 s.f., whichever is smaller.

	M_n	N	β	γ	σ
Central					
$z = 0.1$	11.95	0.032	1.61	0.54	0.11
Total					
$z = 0.1$	11.89	0.031	1.77	0.52	0.10
Evolution					
$z > 0.1$	0.4	− 0.02	− 0.6	− 0.1	N/A

as there is little evolution at late times. However, if the HMF or SMF were to be strongly evolving at low redshift, this approach would yield a result inconsistent with matching the central population. We further argue firmer constraints are set using the central SMHM relation, where there is higher confidence in the populations being matched.

3.2.2 Redshift evolution $z > 0.1$

We assume that at any redshift satellite galaxies at infall follow the same scaling relations as centrals of equal parent halo mass but then may evolve independently afterwards. Therefore, we require an SMHM relation calibrated at all redshifts to initialize satellites in subhaloes at infall.

Keeping fixed the $z = 0.1$ parameters, the redshift-dependent parameters in equation (5) are fit with a similar iterative method. A complication arises as the high redshift data do not distinguish central and satellite SMF. We therefore use a multiepoch[8] approach without satellite evolution to fit the evolution parameters. We continue to generate central SMF as above then add a satellite SMF and compare to the total SMF from the data minimizing the sum of the root mean squares over eight redshift steps. We generate the satellite SMF[9] from the model for each set of parameters in our parameter space. In Appendix A, we show the result of our iterative method applied to the Davidzon et al. (2017) and SDSS SMFs.

It is important to realize at this stage that, as demonstrated in Fig. 5, the vast majority (>60–80 per cent) of $z = 0.1$ satellites are accreted below $z < 1.0$ irrespective of the exact dynamical time-scale adopted.[10] In this redshift range, the evolution of the SMHM relation has been demonstrated to be rather weak (e.g. Moster et al. 2013; Behroozi et al. 2013; Shankar et al. 2014a; Bernardi et al. 2016, and references thereof). Thus, we expect that our results are robust against possible variations in the redshift dependence of the SMHM relation used to initialize infalling satellites, as demonstrated below.

We will now focus on specifically reproducing the local SSMF (red squares in the right-hand panel of Fig. 4).

3.3 Satellite evolution after infall

After infall, satellites are expected to evolve their stellar mass and structure through several processes such as star formation, stellar stripping, and quenching. In this section, we briefly describe each of these processes and the way we have modelled them in this work. We will then discuss in Section 4.2 how these processes may impact, if at all, the predicted local abundances of satellite galaxies.

3.3.1 Star formation rates

We use the SFR parametrization from Lee et al. (2015) with parameters[11] from Tomczak et al. (2016), where s_0 and M_0 have units $\log(\mathrm{M}_\odot)$ and $\mathrm{M}_\odot$, respectively,

$$\begin{aligned}\log[\psi(z, M_*)] &= s_0(z) - \log\left[1 + \left(\frac{M_*}{M_0(z)}\right)^{-\alpha(z)}\right],\\ s_0(z) &= 0.195 + 1.157z - 0.143(z^2),\\ \log[M_0(z)] &= 9.244 + 0.753z - 0.090(z^2),\\ \alpha(z) &= Y1.118.\end{aligned} \tag{6}$$

As discussed below, we will also need to explore different recipes for SFRs. To this purpose, following a continuity-equation approach (Leja et al. 2015), we have refitted the parameters in equation (6) to match the SFR distributions expected from self-consistently growing the SSMF in our models (for further details on this, see Ap-

[9]Under the assumption that satellites are 'frozen' in stellar mass after infall. We will discuss the impact of relaxing this assumption in Section 3.3.

[10]We note this does not imply an absence of satellite galaxies at higher redshift, only that these satellites at higher redshift mostly do not survive to be observed at $z = 0.1$.

[11]These parameters are derived by fitting data from ZFORGE in combination with far-IR imaging from *Spitzer* and *Herschel* in the range of $0.5 < z < 4$. In this work, we extrapolate their fits down to $z = 0$, as this is consistent with the SFR measured by Salim et al. (2007) at lower redshifts.

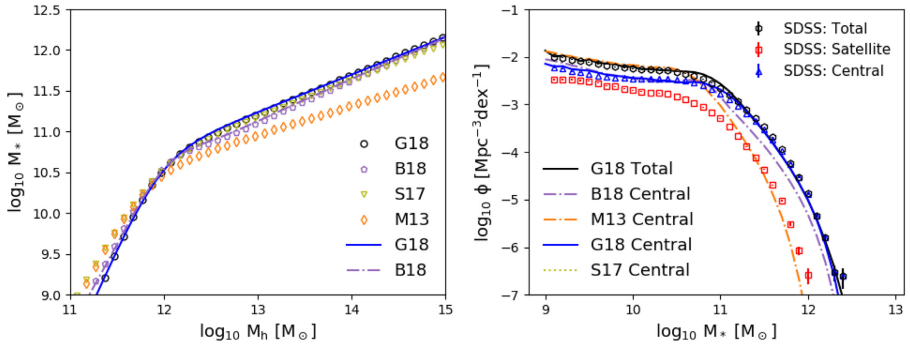


Figure 4. Left-hand panel: The SMHM relation generated from this work (G18, black circles), Behroozi et al. (2018, B18), Shankar et al. (2017a, S17) and Moster et al. (2013, M13). Relations derived from the total SMF/HMF are shown with open symbols, whilst relations derived with the central SMF/HMF are shown as solid lines. Right-hand panel: Central SMF generated from the models using a central only fit, we also show the total SMF from this work (solid black line) from a single-epoch fit. Open symbols show the SDSS central (squares), satellite (triangles), and Total (hexagons) SMFs.

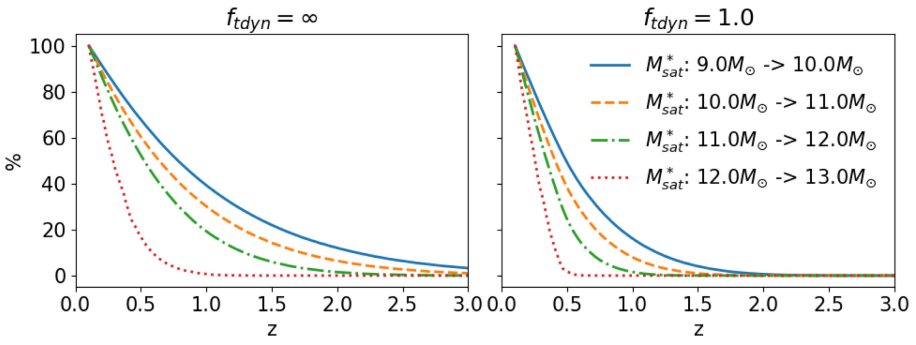


Figure 5. We here show the percentage of satellites observed at a $z = 0.1$ as a function of their redshift of accretion $z > 0$. It can be seen that massive satellites observed at $z = 0.1$ are accreted more recently than smaller satellites. At $z = 0.5$ less that 50 per cent of the total satellites observed at $z = 0.1$ have been accreted, and at $z = 0.1$ this falls to less than 20 per cent.

pendix B). The fit to the resulting SFRs is given by

$$\log(\psi(z, M_*)) = s_0(z) - \log\left[1 + \left(\frac{M_*}{M_0(z)}\right)^{-\alpha(z)}\right],$$
$$s_0(z) = 0.6 + 1.22z - 0.2(z^2),$$
$$\log(M_0(z)) = 10.3 + 0.753z - 0.15(z^2),$$
$$\alpha(z) = 1.3 - 0.1z. \quad (7)$$

In all cases, the SFR is included in our models with a lognormal scatter of 0.3 dex (Leja et al. 2015).

3.3.2 *Quenching*

We also include the ability to quench star formation in satellite galaxies after infall. It has been suggested that satellites undergo a 'delayed-then-rapid' quenching (Wetzel et al. 2013). The latter model envisions that satellites continue to form stars at the same rate as central galaxies of comparable stellar mass for a time τ_q after infall, and then quench rapidly over a time-scale τ_f. This quenching is proposed for satellites with stellar mass above $M_* \gtrsim 10^9 M_\odot$, with a minimum $\tau_q \sim 1$ Gyr. For galaxies below $M_* \lesssim 10^9 M_\odot$, we however adopt the more recent results by Fillingham et al. (2016), who put forwards a parent halo mass dependent cut-off,

$$\log(M_{\rm cutoff}) = 9\log(M_\odot) - (15\log(M_\odot) - \log(M_{\rm h,host}))/5\log(M_\odot), \quad (8)$$

below which satellite galaxies all share the same quenching time $\tau_q = 2$Gyr. The rapid quenching time-scale τ_f can be expressed as (Wetzel et al. 2013)

$$\tau_f = -0.5\log(M_{*,\rm sat}) + 5.7\,\text{Gyr}. \quad (9)$$

We set a minimum τ_f of 0.2 Gyr for all galaxies. This rapid quenching begins at times $t > \tau_q$ after infall, and it is approximated by an exponential decay that is longer for larger satellites, as can be inferred from equation (10). The lookback time at which a galaxy begins fast quenching is then $t_q = t_{infall} - \tau_q$. After this time the

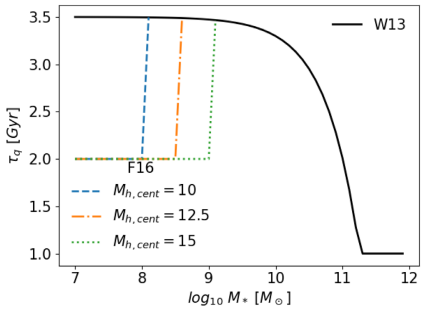


Figure 6. The solid line shows the Wetzel et al. (2013, W13) model for quenching. The dashed lines show the host halo-dependent reduction in quenching time from Fillingham et al. (2016, F16) for three example host masses $\log 10\, M_{\rm h,cent} = 10$, 12.5, 15 as labelled. Larger hosts are able to reduce the quenching time of larger satellites.

satellite no longer follows the SFR of a typical central galaxy. Fig. 6 illustrates the quenching model, where the dashed/coloured lines demonstrate the halo mass dependence in cut-off mass.

The SFR during the satellite infall is then given by

$$\mathrm{SFR}(t, M_*) = \mathrm{SFR}(t, M_*) \begin{cases} \psi(z(t), M_*), & t > t_q \\ \psi(z(t_q), M_*) \mathrm{e}^{\left[-\frac{t_q - t}{\tau_f}\right]}. & t < t_q \end{cases} \quad (10)$$

If at any point a satellite galaxy has an SSFR below 10^{-12} $\mathrm{M}_\odot$ yr^{-1}, it is assumed to be fully quenched and assigned an SSFR of 10^{-12} $\mathrm{M}_\odot$ yr^{-1}, plus a lognormal scatter of 0.3 dex.

3.3.3 Mass recycling

During infall some of the new stellar mass formed may be returned to the interstellar medium after stellar death. To incorporate this mass recycling process, we use the fractional mass-loss as parametrized in Moster, Naab & White (2018)

$$f(\tau_{\rm ml}) = 0.05 \ln\left(\frac{\tau_{\rm ml}}{1.4\mathrm{Myr}} + 1\right). \quad (11)$$

It follows that the mass-loss rate (MLR) at a given time is dependent on the star formation history (SFH), and it is calculated as

$$\mathrm{MLR}(t) = \frac{\sum_{t'=t_{\rm infall}}^{t} \mathrm{SFH}(t')(f[t' - (t - \delta t)] - f[t' - t])}{\delta t}. \quad (12)$$

In equation (12), the SFH is the mass of stars formed in each previous time-step. The actual stellar mass growth $\dot{M}_{*,\rm sat}$ is then given by the difference between the SFR and the MLR

$$\dot{M}_{*,\rm sat}(t) = \mathrm{SFR}(M_{*,\rm sat}, t) - \mathrm{MLR}(t). \quad (13)$$

3.3.4 Stripping

The stellar stripping is implemented following the empirically based formalism suggested by Cattaneo et al. (2011). Satellite galaxies strip stellar mass proportionally to the ratio of the host subhalo and the parent halo

$$M_{*,\rm sat} = M_{*,\rm sat}(1 - \eta_{\rm strip})^{\tau_{\rm strip}}, \quad (14)$$

where $\tau_{\rm strip}$ is the dynamical friction time-scale in orbital time. For our model, where the orbital circularity is assumed to take the average value, this becomes the dynamical friction time (Chandrasekhar 1943) times a constant given by Jiang et al. (2008) using the average orbital circularity. Following Cattaneo et al. (2011), the latter can be written

$$\tau_{\rm strip} = \frac{1.428}{2\pi} \frac{M_{\rm h,cent}/M_{\rm h,sat}}{\ln(1 + M_{\rm h,cent}/M_{\rm h,sat})}. \quad (15)$$

We set as a reference $\eta_{\rm strip} = 0.4$ as suggested by Cattaneo et al. (2011) who showed that larger values would be inconsistent with the observational constraints in the local Universe. If at the time of observation, say $z = 0$, a galaxy's full dynamical time is not yet passed, a time-dependent reduction factor is applied to the amount of total stellar stripping given by equation (15).

4 NUMBER DENSITIES AND DISTRIBUTIONS OF MASSIVE SATELLITES

Section 3 has laid out our methodology, which is essentially based on progressively building up the (surviving) satellite number densities over cosmic time based on the input dynamical merger time-scales. Our first goal is now to predict the local ($z = 0.1$) satellite SMF and the distribution of massive satellites as a function of parent halo mass, which we will compare with our refined SDSS galaxy-halo catalogue introduced in Section 2. When computing the distribution of satellites as a function of parent halo mass we will show both full number densities, as well as fractional distributions to better highlight the 'skewness' of the predicted distributions with respect to the data. The latter will be simply computed as

$$F(\mathrm{d}M_{\rm h}) = \frac{N(>10^x)|_{\mathrm{d}M_{\rm h}}}{N(x)}, \quad (16)$$

where $N(>10^x)$ is the total number (density) of satellites above a threshold stellar mass $x = \log M_*$, and the $N(>10^x)|_{\mathrm{d}M_{\rm h}}$ is the number of these that reside within the halo mass bin $[M_{\rm h}, M_{\rm h} + \mathrm{d}M_{\rm h}]$.

The results discussed in the next sections rely heavily on the concept of mass ratio affecting the merging time-scales of subhaloes and/or satellite galaxies. To ease the discussion of our findings, we provide Fig. 7 as visual guide to how merging times depend on subhalo/satellite masses in three different parent halo masses. In the left-hand panel, we plot dynamical time against subhalo mass for a subhalo accreted at $z = 1.5$. The black dotted line is the time (in Gyr) to $z = 0$. If a subhalo's $\tau_{\rm merge}$ is below this value, it will not be observed at redshift $z = 0$. Similarly, for the right-hand panel, we plot the same relation converting the subhalo masses to satellite galaxy masses using the SMHM relation at $z = 1.5$. In both panels, the merging time is increased when the ratio of mass between the subhalo/satellite and the central (parent) halo is smaller.

In Section 4.1, we will first focus on the frozen model. In Section 4.1.1, we will motivate the need for a finite dynamical time and an evolving SMHM relation, and then in Section 4.1.2 set constraints on the dynamical times that well reproduce the number and distribution of SDSS satellites as a function of parent host halo mass. In Section 4.2, we will probe the impact of SFR, stripping and quenching in the predicted satellite distribution in the local Universe.

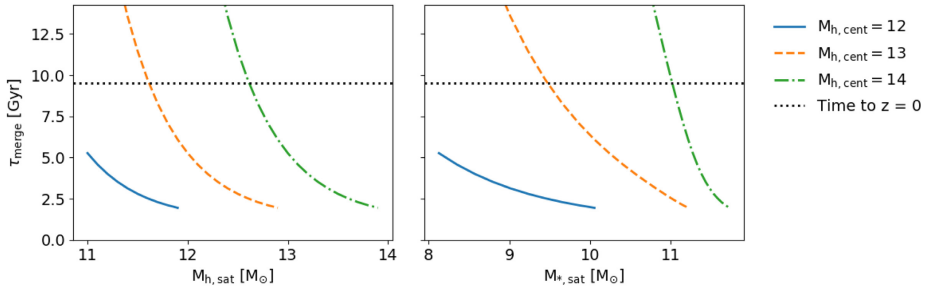


Figure 7. Range of merging time-scales for a range of subhalo (left) and satellite (right) masses when accreted on to three different host masses: log $M_{\rm h,cen}$ [$M_\odot$] = 12 (blue solid), 13 (orange dashed), and 14 (green dot–dashed). The dotted black line shows the time to redshift $z = 0$, i.e. the minimum amount of time a satellite would need to survive to be observable in the local Universe.

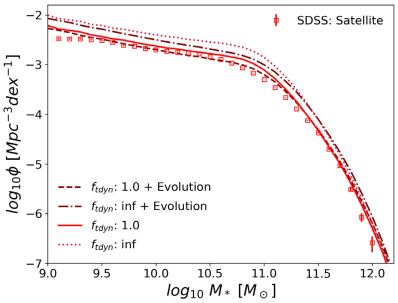


Figure 8. SSMFs generated by the model compared to SDSS data (open squares). The dashed and solid lines refer to models with a reference $f_{\rm tdyn} = 1.0$ (as suggested by N-body simulations), respectively, with and without redshift evolution in the input SMHM relation. The dot–dashed and dotted lines have infinite $f_{\rm tdyn}$ (i.e. no mergers) with and without redshift evolution in the input SMHM relation.

4.1 Frozen model

We begin by showing the model predictions of satellite number densities and distributions in parent halo mass in a frozen model, where satellites do not evolve in stellar mass after infall.

4.1.1 Redshift evolution and mergers

We start by exploring a basic model where $\tau_{\rm dyn}$ is infinitely long (no satellite mergers) and the SMHM relation does not evolve in cosmic time, i.e. keeping the same $z = 0$ parameters at all redshifts. We choose to start with the simplest assumption of an unevolving SMHM, as also suggested by some empirical evidence at least up to $z \sim 1$ (e.g. Shankar et al. 2006, 2014b; Moster et al. 2013; Buchan & Shankar 2016; Tinker et al. 2017; Guo et al. 2018). We then compare the outputs of this model with those from models in which we either allow for galaxy–satellite mergers ($f_{t_{\rm dyn}}$ is finite), and/or the SMHM relation evolves with redshift. Fig. 8 shows how the four combinations of these assumptions affect the satellite SMF. For both the models characterized by an infinite dynamical time ($f_{\rm tdyn} = \infty$, dot–dashed and dotted lines), the number densities of satellites are too high compared to the observed satellite SMF (red squares). When setting $f_{\rm tdyn} = 1.0$, i.e. the merger time-scales are exactly those parametrized by McCavana et al. (2012), both the evolving and unevolving SMHM models produce very similar number densities and well consistent with the observed satellite SMF. The model with an evolving SMHM relation is particularly successful.

We now test in Fig. 9 the four models against the SDSS distributions of the satellites in groups and clusters. The top row shows the comoving number density per halo mass bin. The bottom row shows the fractional distributions (equation 16) that better highlights the differences induced by varying model inputs. In both rows, the SDSS data are shown as grey bands. As expected, the largest difference is seen between the models with $f_{\rm tdyn} = \infty$ and $f_{\rm tdyn} = 1.0$. When $\tau_{\rm dyn}$ is infinite,[12] the distribution of satellites skews such that high-mass satellites are found more frequently in relatively lower mass parent haloes. We explain the origin of this effect when discussing dynamical time in more detail in Section 4.1.2. The evolution[13] of the SMHM relation has a low impact on the number densities. This is in part due to the evolution of the SMHM relation being weak especially in the high-mass end, and in part due to most (>60–80 per cent $z = 0.1$) satellites being accreted at relatively low redshift ($z < 1.0$). We thus support the view that local satellite distributions are largely independent of the specific input of the evolution parameters of the SMHM relation.

4.1.2 Dynamical time

In this section, we show the predictions of models with different merging time-scales, $f_{\rm tdyn} = 0.5, 1.0, 2.5$, to probe the effects of dynamical time on the satellite population. In Fig. 10, we find, as expected, that longer dynamical times tend to increase satellite number densities especially towards lower stellar masses. The number densities of higher-mass satellites are more resilient to increase with dynamical time. In fact, when $f_{\rm tdyn} \gtrsim 1$ the number densities of massive satellites become already very close to the theoretical maximum number density (Fig. 8) set by $f_{t_{\rm dyn}} = \infty$. It follows that massive satellites are on average a recently accreted population. In

[12] In PYTHON in this case, we set this variable using 'numpy.inf'.
[13] We direct readers to Appendix A for further discussion and motivation for the redshift evolution parameters used in this work.

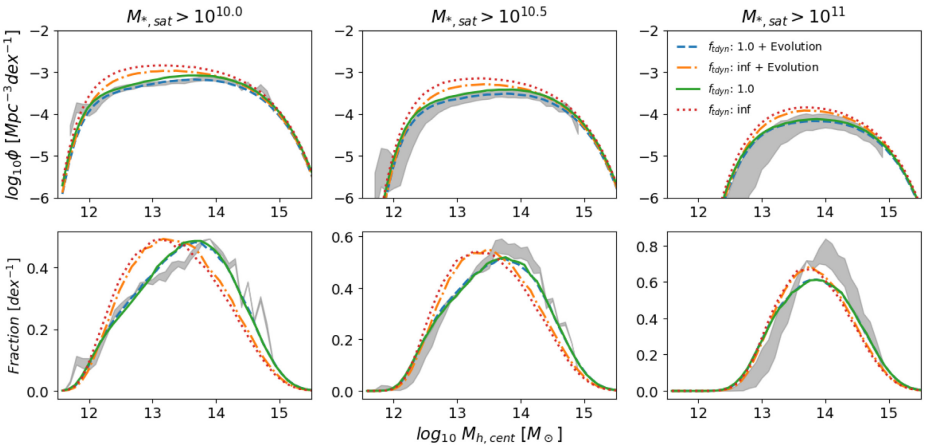


Figure 9. Satellite distributions in parent haloes generated from the model are compared to those observed in SDSS (grey band). Columns from left to right show increasing satellite stellar mass cuts as labelled. The top row shows the number density of satellites expected to be found in each parent halo mass. The bottom row shows the fractional distribution described by equation (16). The solid and dashed lines show a reference $f_{\rm rdyn} = 1.0$ with and without evolution in the SMHM relation, respectively. The dot–dashed and dotted lines show $f_{\rm rdyn} = \infty$ (i.e. no mergers) with and without evolution in the SMHM relation, respectively. The width of the grey band corresponds to a 10 per cent uncertainty in satellite stellar masses.

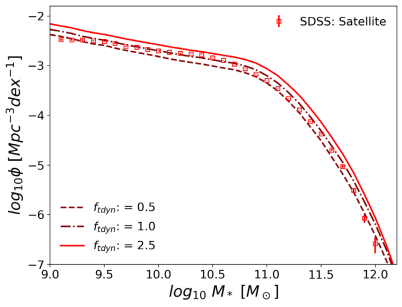


Figure 10. SSMFs generated by the model compared to SDSS data (open squares). The solid, dot–dashed, and dashed lines show $f_{\rm rdyn} = 0.5$, 1.0, and 2.5, respectively.

other words, there are only a few high mass satellites that had enough time to merge when $f_{\rm rdyn} \gtrsim 1$. In contrast, lower mass satellites have not yet reached their theoretical limit, and thus they still have room to increase their number densities with increasing dynamical time.

Similarly to Fig. 9, Fig. 11 shows how the distribution of satellites as a function of parent halo mass is affected by $f_{\rm rdyn}$. The number density distribution (top row) shows results similar to the satellite SMF where increasing dynamical time increases the number densities. However, there is also an apparent steepening effect for which lower mass host haloes end up containing relatively less satellites with respect to models with longer dynamical times. The fractional plot (bottom row) accentuates this change in the number density distributions shown in the top row: shorter dynamical times shift the peak of the distribution to the right as relatively more satellites are observed in high-mass host haloes.

This steepening of the satellite distribution as a function of halo mass, as well as the shift mentioned in Section 4.1.1 where infinite dynamical times move satellites preferentially to lower halo masses, are both caused by the amount of time satellites survive in their hosts. Massive satellites are far more common in massive hosts, as can be inferred from the SHMF. Therefore, irrespective of the chosen merging time-scale, there will always be a high number density of surviving massive satellites in higher mass parent haloes. However, when merging time-scales are increased, the lower number densities of massive satellites in moderately-sized haloes are also increased. Given that lower mass parent haloes are more abundant, the reduction of merging time-scales tends to shift the peak of the fractional distribution of galaxies to lower mass parents. Otherwise put, the merging time-scales in clusters ($\log_{10} M_{\rm h,cent}$ [$M_\odot$] >13.75) are so long that even a factor of 5 reduction in merger time-scale still does not give the satellite galaxies sufficient time to merge with the central galaxy. This effect can be inferred from Fig. 7, where we see steeper gradients for a given subhalo/satellite mass with increasing parent halo mass. Merging is more efficient for the lower mass satellites, as can be seen by the steepness of the $f_{\rm rdyn} = 0.5$ model (dashed lines) in the top row of Fig. 11.

The least-square residuals to the SMF, number density distribution and fractional distributions are given in Table 3. There is no model that simultaneously fully matches all the observations in all mass ranges. The trend in both the number density distribution and the fractional distribution is that slightly longer dynamical times ($f_{\rm rdyn} = 1.2$) are favoured by the less massive satellites. Longer dynamical time-scales better match the halo mass distributions (number density distribution/fractional distributions) for lower mass satellites, and vice versa for higher mass satellites with $\log M_*/M_\odot > 11$. Nevertheless, the simple combination of abundance matching

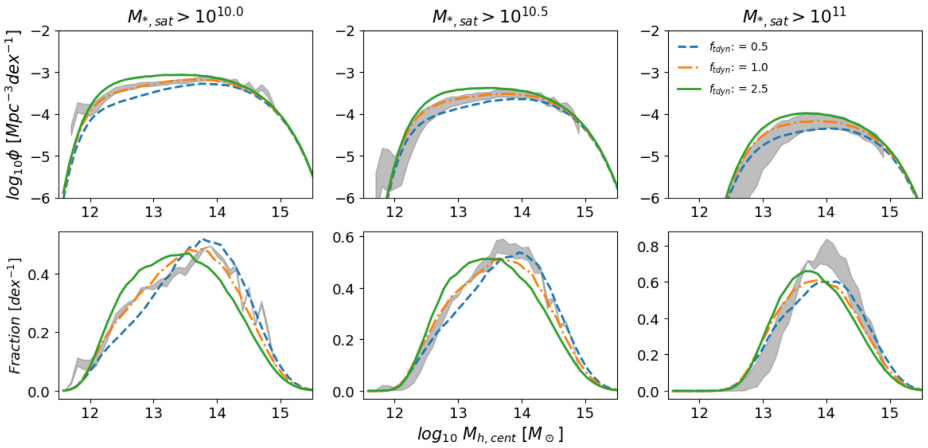


Figure 11. Satellite distributions in parent haloes generated from the model are compared to those observed in SDSS (grey band). Columns from left to right show increasing satellite stellar mass cuts as labelled. The top row shows the number density of satellites expected to be found in each parent halo mass. The bottom row shows the fractional distribution described by equation (16). The dashed, dot–dashed, and solid lines show $f_{tdyn} = 0.5$, 1.0, and 2.5, respectively. The width of the grey band corresponds to a 10 per cent uncertainty in satellite stellar masses.

Table 3. We show the sum of the squared residuals between the SDSS and our model. The satellite SMF is calculated between 9.1 and 12.0 M_*. The SDF fit is calculated between 12 and 14.9 M_h for the >10 and >10.5 plots, and between 12.5 and 14.9 M_h for >11. The fractional plot fit is calculated between 11.6 and 14.9 M_h.

f_{tdyn}	SSMF (Fig. 10) (Fig. 10)	SDF (Top row Fig. 11)			Fractional distribution (Bottom row Fig. 11)		
		>10	>10.5	>11	>10	>10.5	>11
0.5	0.022	0.19	0.55	0.073	0.0042	0.0047	0.0078
0.8	0.025	0.13	0.51	0.089	0.0020	0.0017	0.0054
1.0	0.034	0.12	0.56	0.10	0.0015	0.0011	0.0050
1.2	0.043	0.12	0.53	0.11	0.0015	0.00094	0.0046
1.5	0.054	0.12	0.52	0.13	0.0017	0.0010	0.0045

and dynamical merging time-scales as suggested by pure *N*-body simulations ($f_{tdyn} = 1.0$) tends to provide overall good agreement to both the satellite SMF and the satellite distributions, without the need to invoke additional physics in the (late) evolution of satellite galaxies after infall.

4.2 'Non-frozen' evolutionary models

We showed in the Section 4.1 that the number and distribution of satellite galaxies are primarily driven by the accretion history and dynamical friction time-scale, and depend weakly on the exact redshift evolution in the input SMHM relation.

However, as discussed in Section 3.3, satellites may not be a strictly frozen population. After infall a number of processes such as star formation, quenching and/or stellar stripping may all well affect the final stellar masses of satellite galaxies. In this section, we explore the impact that each one of these processes could have on the predicted local population of satellite galaxies. We will prove, in particular, that additional, late stellar stripping, star-formation, and quenching, whilst not necessarily playing a major role in varying integrated number densities, are still all necessary ingredients to fully reproduce SDSS data in terms of, for example, specific SFR distributions.

In what follows, we explore the two SFR models with the quenching model described in Sections 3.3.1 and 3.3.2. The first, purely observationally-based, SFR model strictly follows the SFR parametrization by Tomczak et al. (2016, T16 hereafter), given in equation (6). The second SFR model (which we label as 'CE") is instead based on a continuity equation approach similar to Leja et al. (2015). In essence, the latter is based on first assuming number conservation of galaxies, and then deriving the implied SFRs as a function of stellar mass from the time growth in any stellar mass bin implied by the redshift evolution of the SMF (corrected for gas loss from dying stars assumed to be an average of 40 per cent). The novelty in the latter model with respect to previous work is that we do not tune the resulting SFR on the total but rather only on the SMF of *central* galaxies, which is in turn iteratively constrained by matching the local SMF of SDSS centrals. This has the main effect of lowering the implied average SFR at fixed stellar mass at any given epoch (see equations 6 and 7).

We compare the satellite SMFs produced by the two star formation+quenching models addressed above to our SDSS SMF

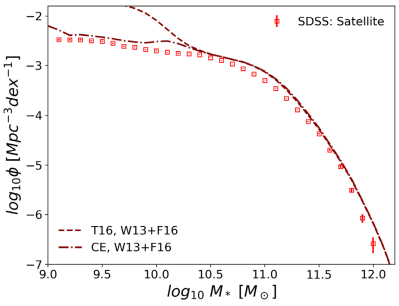


Figure 12. SSMFs generated from the model using both the Tomczak et al. (2016) (dashed) and continuity (dot–dashed) SFRs compared to the SDSS SSMF (open squares).

of satellites in Fig. 12. It is apparent that using the observed SFR by T16 (dashed line), even inclusive of the best recipes for quenching, still substantially overproduces the number density of galaxies below $M_* \lesssim 3 \times 10^{10}\,\mathrm{M}_\odot$. This is a well-known problem affecting the full (dominated by central) galaxy population (e.g. Leja et al. 2015): the integrated (observed) SFR is not consistent with the moderate growth over time of the SMF causing an overproduction of galaxies becoming gradually more severe at lower stellar masses. Our results point to a similar problem affecting the satellite population, on the assumption that the latter at infall share the same SFR distribution as a typical central galaxy of the same stellar mass.

To further constrain our adopted continuity equation-based SFR distributions, we compare our best models with the bimodality in (satellite) SSFR measured in the SDSS data (grey shading in Fig. 13). The data show a clear bimodality in SSFR at least for galaxies below $M_* \lesssim 10^{11}\,\mathrm{M}_\odot$, at variance with the SSFR distribution of centrals (black thick long dashed lines), which point to a much larger population of star-forming galaxies. This difference in SSFR distributions has been often ascribed to environmental quenching (e.g. Peng et al. 2010), either in the form of stripping or strangulation. Both the SFR reference models predict very similar distributions in SSFR for galaxies above $M_* \gtrsim 3 \times 10^{10}\,\mathrm{M}_\odot$, as expected from the invariance in the predicted number densities above this stellar mass. When using the continuity equation-based method (CE, dot–dashed line) the overproduction is greatly reduced and the SSMF is very close to the SSMF with no evolution in the satellite population.

We now move on showing the relative impact of SFR and stellar stripping on the SSMF. Figs 14 and 15 show SMF and host halo mass distributions for the $f_{t_{\rm dyn}} = 1.0$ reference model with neither stripping nor star formation, the CE star formation model, and the CE star formation model with stripping (long-dashed, dot–dashed, and solid lines, respectively). We see from both figures that the reference and star formation model are almost indistinguishable. The stripping, at least at the level implemented in this work, also has a rather minor effect, at the most reducing the number densities of the most massive satellites ($>10^{11}\mathrm{M}_\odot$) by $\lesssim 0.2$ dex. Note that that the *cumulative* effect of stripping on the stellar mass growth of central galaxies could be much more prominent, as emphasized by a number of groups (e.g. Cattaneo et al. 2017).

Table 4 shows the sum of the square residuals to the SMF, SDF, and fractional distributions for the same models discussed above, our reference frozen one, and the one with evolution of satellites after infall (stellar stripping and continuity equation-based star formation). Table 4 shows that the satellite late evolution has little effect on the fractional distribution. In the number density distribution, we see an improved fit for galaxies in the $M_* > 10.^{11}\,\mathrm{M}_\odot$ range, mainly induced by the stripping which slightly reduces the number density of massive galaxies. Table 4 shows the sum of square residuals for the dynamical time with $f_{t_{\rm dyn}} = 1$, for the frozen and evolved models.

In summary, our results point to the fact that direct abundance matching between centrals and satellite galaxies and the halo plus SHMFs in the local Universe (and at higher redshifts) is a good approximation to the mean SMHM relation for central galaxies, irrespective of the precise late evolution of satellites after infall.

5 DISCUSSION

5.1 Strengths of the statistical approach with respect to the state of the art

The statistical semi-empirical model introduced in this work presents a highly complementary approach to the existing suit of cosmological galaxy evolution models. The latter models, either being fully analytic or fully numerical, inevitably rely on large boxes to simulate both low- and high-mass galaxies with sufficient statistics. Even traditional semi-empirical models, based on applying abundance matching techniques to vast catalogues of dark matter merger trees (e.g. Behroozi et al. 2018, and references therein), still require large computational resources. Our approach, based on statistical mean dark matter accretion histories, allows to explore the full range of galaxy stellar and halo masses without the need to simulate large volumes or even weight dark matter merger trees. Whilst we lose the ability to track an individual galaxy through cosmic time, we are able to rapidly predict the statistics of any subpopulation of chosen galaxies at any redshift and environment.

There is vast literature on the modelling of satellite galaxies. Here, we recall just a couple of examples of semi-empirical and semi-analytic models to highlight some of the key similarities and key differences. Neistein et al. (2013), with an approach similar to ours, separated the central and satellite populations in an attempt to better define the galaxy halo connection. By allowing in an *N*-body simulation the stellar mass of satellite halaxies to depend on both the host subhalo mass and on the parent halo mass, Neistein et al. (2013) find that the local satellite SMHM is substantially less well defined than the one for central galaxies. In our model, satellites instead strictly follow the SMHM relation of centrals at infall. In this way, we find the resulting satellite distributions to be well reproduced. Our semi-empirical statistical model was able to reproduce multiple observables such as the stellar and parent halo mass distributions, with essentially only one parameter, $f_{t_{\rm dyn}}$. By working with minimal assumptions and related free parameters, our approach is thus less prone to possible degeneracies affecting more traditional, multiparameter techniques.

Another key difference with respect to previous models concerns 'orphan galaxies'. In *N*-body or merger tree-based simulations, when a subhalo goes below the resolution limit, an orphan galaxy is created (e.g. De Lucia et al. 2011; Guo et al. 2011). It is then necessary to make an assumption on how much longer that subhalo (and hosted satellite galaxy) will survive. In our model, we avoid this complication by self-consistently assigning to all satellites a (full) observability time-scale at infall.

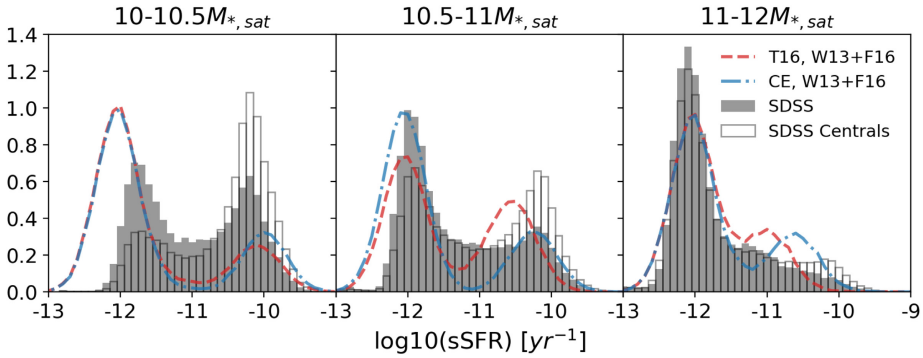


Figure 13. Specific SFR distributions in three mass ranges (as labelled) generated from the model using both the Tomczak et al. (2016) (dashed) and continuity (dot–dashed) SFRs. Data shown are the distribution of specific SFR in SDSS satellites (shaded bars) and SDSS centrals (unfilled bars).

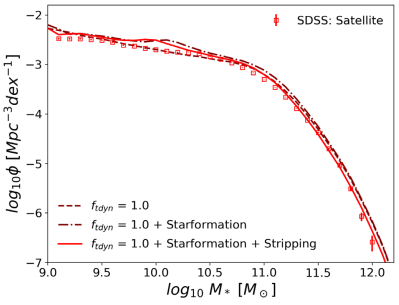


Figure 14. SSMFs generated from the model compared to SDSS satellites (open squares). The models shown all have $f_{tdyn} = 1.0$ and are the reference 'frozen model' (dashed line), starformation (CE model) only (dot–dashed line) and star formation and stripping (solid line).

5.2 Future applications of the statistical semi-empirical model

We have shown here the ability of the statistical semi-empirical model to reproduce the observed group and cluster richness at $z = 0.1$ using simply the dynamical friction time-scale from *N*-body simulations. In future work, we will also compare our model outputs with the richness of groups and clusters at high redshift thus providing additional important constraints to models of structure formation. For example, we will use our statistical modelling to fit extreme objects such as the high redshift ($z = 2.5$) massive cluster reported in Wang et al. (2016), which has been claimed to be challenging in other ΛCDM cosmology-based models. By predicting the (correct) distributions of satellite galaxies predicted over many epochs, more accurate merger rates and pair fractions can be calculated and compared to observations found in surveys such as those reported in Mundy et al. (2017). Via more robust estimates of galaxy merger rate, we will be able to set upper limits on the impact of mergers in shaping galaxy morphology (Hopkins et al. 2009a), build galactic bulges (Cole et al. 2000; Hopkins et al. 2010; Shankar et al. 2013), and trigger AGN activity (Villforth et al. 2017).

Our technique being fast and flexible is an extremely valuable resource in view of the next large and deep extra-galactic surveys for example *Euclid* (Refregier et al. 2010). Our model can be both predictive, working out specific expectations for such surveys, and then reactive, using the model outputs as constraints on existing formation models.

6 CONCLUSIONS

In this paper, we have presented 'STEEL' a *STatistical* sEmi-Empirical modeL that replaces the traditional merger tree dark matter backbone with a 'statistical accretion history'. In essence, we first trace backwards the mean accretion history of dark matter haloes in a given bin of halo mass at the redshift of interest z. We then gradually build in time their satellite population by integration of the SHMF, and at each time-step assign stellar masses to subhaloes via abundance matching techniques. Our approach is extremely fast and accurate, allowing to probe the galaxy number densities and average properties within the full range of stellar masses probed by large extragalactic surveys, with virtually no limitation on volume size or mass resolution.

We use the statistical semi-empirical model to predict the satellite galaxy population at stellar masses above $M_* \gtrsim 10^9\,M_\odot$. We find that irrespective of the exact input SMHM relation, the main driver shaping the local SMF of satellite galaxies is by far the input dynamical friction time-scale. In particular, adopting a merging time-scale very close to the one calibrated from high-resolution *N*-body dark matter simulations provides an excellent match to the satellites SMF as inferred from SDSS. Shorter dynamical friction time-scales not only reduce the number densities of satellites (as more satellites disappear merging with central galaxies) but also shift the abundances of satellites into higher mass parent haloes than actually observed. Conversely, longer dynamical friction time-scales create an overproduction in the SSMF and induce too many satellites in lower mass parents. Thus, a traditional hierarchical dark matter cosmology naturally provides the right abundances of satellites in the local Universe without any additional fine-tuning.

We then incorporate popular recipes for star formation, stellar stripping, and quenching in satellite galaxies after infall. We find

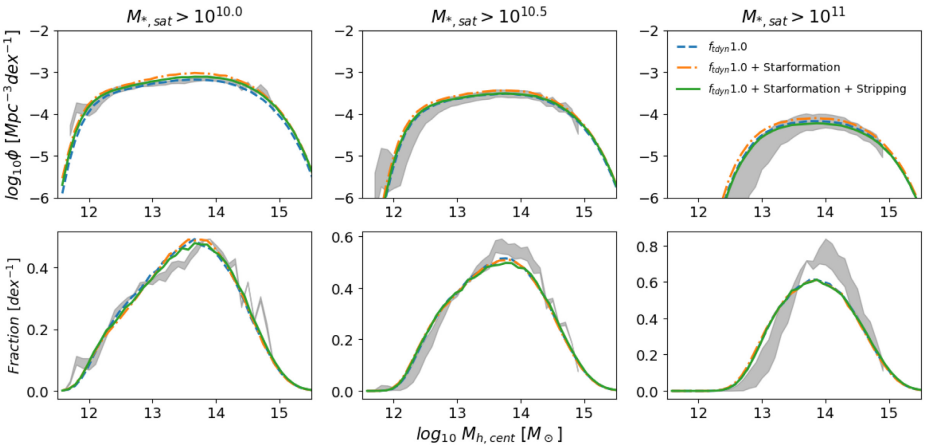


Figure 15. Satellite distributions in parent haloes generated from the model are compared to those observed in SDSS (grey band). Columns from left to right show increasing satellite stellar mass cuts as labelled. The top row shows the number density of satellites expected to be found in each parent halo mass. The bottom row shows the fractional distribution described by equation (16). The models shown all have $f_{tdyn} = 1.0$ and are the reference 'frozen model' (dashed line), starformation only (dot–dashed line) and star formation and stripping (solid line). The width of the grey band corresponds to a 10 per cent uncertainty in satellite stellar masses.

Table 4. We show the sum of the squared residuals between the SDSS and our model as in Table 3 with the same mass ranges for the fitting. All models have $f_{t_{dyn}} = 1.0$ from top to bottom, we then have the reference frozen model, the model with starformation, and the model with stripping and star formation.

$f_{t_{dyn}}$	SSMF (Fig. 14)	SDF (Top row Fig. 15)			Fractional distribution (Bottom row Fig. 15)		
		>10	>10.5	>11	>10	>10.5	>11
1.0	0.034	0.12	0.53	0.10	0.0015	0.0011	0.0049
1.0							
With star formation	0.049	0.077	0.43	0.14	0.0018	0.0012	0.0059
1.0							
With stripping and star formation	0.021	0.087	0.47	0.088	0.0016	0.0015	0.0056

that, within reasonable variations of the input parameters, all these processes play a secondary role. At fixed merging time-scale the resulting SMF and halo mass distribution of local satellites end up being very similar. All in all, our results highly suggest that at any given epoch, a direct abundance matching between the total galaxy SMF and halo plus SHMFs provides a very good approximation to the mean SMHM relation of central galaxies, irrespective of the exact (late) evolution of satellites after infall. In other words, our results confirm the common assumption that at any given epoch satellites can be abundance matched as centrals at infall, irrespective of their redshift of infall.

We find that, similarly to what previously noticed for centrals, adopting the latest empirical determinations of the SFR as a function of stellar mass and redshift significantly overproduces the abundance of galaxies below $M_* \lesssim 3 \times 10^{10}\,M_\odot$, irrespective of the quenching, stripping, or merger time-scales adopted. An SFR as predicted by a continuity equation approach well matches both the local abundances of satellites and the bimodality in specific SFRs for all galaxies above $M_* \gtrsim 10^{11}\,M_\odot$ as measured in the SDSS.

STEEL is an ideal tool to rapidly compute accurate galaxy merger rates as a function of stellar mass, time, and environment. It can reveal the impact of mergers in shaping the structural and dynamical evolution of galaxies, and eventually their central supermassive black holes. We will look into some of these topics in more detail in future work.

ACKNOWLEDGEMENTS

We warmly thank B. Moster and J. Leja for detailed discussions on the SMHM relation and the SFR, respectively. We also thank C. Conselice, N. Menci, and C. Marsden for useful discussions on the project. We also acknowledge extensive use of the PYTHON libraries ASTROPY, MATPLOTLIB, NUMPY, PANDAS, and SCIPY. PJG acknowledges support from the STFC for funding this PhD.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Baugh C. M., 2006, A Primer on Herarchical Galaxy Formation: The Semi-analytical Approach. Available at: http://iopscience.iop.org/article/10.1086/379875/pdf; http://stacks.iop.org/0004-637X/600/i=2/a = 580
Behroozi P., Wechsler R., Hearin A., Conroy C., 2018, preprint (arXiv: 1806.07893)

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernardi M., Fischer J. L., Sheth R. K., Meert A., Huertas-Company M., Shankar F., Vikram V., 2017b, MNRAS, 468, 2569
Bernardi M., Meert A., Sheth R. K., Fischer J. L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2017a, MNRAS, 467, 2217
Bernardi M., Meert A., Sheth R. K., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016, MNRAS, 455, 4122
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Bosch F. C. v. d., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1
Bosch F. C. v. d., Ogiya G., Hahn O., Burkert A., 2018, MNRAS, 474, 3043
Boylan-Kolchin M., Ma C. P., Quataert E., 2008, MNRAS, 383, 93
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buchan S., Shankar F., 2016, MNRAS, 462, 2001
Cattaneo A., Mamon G. A., Warnick K., Knebe A., 2011, A&A, 533
Cattaneo A. et al., 2017, MNRAS, 471, 1401
Chabrier G., 2003, PASP, 115, 763
Chandrasekhar S., 1943, ApJ, 97, 255
Cole S., Lacey C., Baugh C., Frenk C., 2000, MNRAS, 319, 168
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12
Davidzon I. et al., 2017, A&A, 605, A70
De Lucia G., Fontanot F., Wilman D., Monaco P., 2011, MNRAS, 414, 1439
Diemer B., 2017, Technical report, COLOSSUS: A Python Toolkit for Cosmology, Large-Scale Structure, and Dark Matter Halos, Harvard-Smithsonian Center for Astrophysics. Available at: http://arxiv.org/abs/1712.04512
Fattahi A. et al., 2016, MNRAS, 457, 844
Fillingham S. P., Cooper M. C., Pace A. B., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., Wheeler C., 2016, MNRAS, 463, 1916 (F16)
Giocoli C., Pieri L., Tormen G., 2008, MNRAS, 387, 689
González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, ApJ, 735, L34
Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580
Guo H. et al., 2018, preprint (arXiv:1810.05318)
Guo Q. et al., 2011, MNRAS, 413, 101
Hirschmann M., Naab T., Somerville R. S., Burkert A., Oser L., 2012, MNRAS, 419, 3200
Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009b, ApJ, 691, 1168
Hopkins P. F. et al., 2009a, MNRAS, 397, 802
Hopkins P. F. et al., 2010, ApJ, 715, 202
Huertas-Company M., et al., 2015, ApJ, 809, 95
Huertas-Company M., et al., 2016, MNRAS, 462, 4495
Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, ApJ, 675, 1095
Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848
Khochfar S., Burkert A., 2006, A&A, 445, 403
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlober S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Lapi A., Cavaliere A., 2011, Adv. Astron., 2011, 1
Lee N. et al., 2015, ApJ, 801, 80
Leja J., Van Dokkum P. G., Franx M., Whitaker K. E., 2015, ApJ, 798, 115
Lemson G., Consortium T. V., 2006, Technical Report, Halo and Galaxy Formation Histories from the Millennium Simulation: Public Release of a VO-oriented and SQL-queryable Database for Studying the Evolution of Galaxies in the LambdaCDM Cosmogony. Max Planck Institutes for Astrophysics and for Extraterrestrial Physics
McCavana T., Micic M., Lewis G. F., Sinha M., Sharma S., Holley-Bockelmann K., Bland-Hawthorn J., 2012, MNRAS, 424, 361
Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943
Meert A., Vikram V., Bernardi M., 2016, MNRAS, 455, 2440
Menci N., Fontana A., Giallongo E., Grazian A., Salimbeni S., 2006, ApJ, 647, 753
Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3
Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822
Moster B. P., Somerville R. S., Maulbetsch C., Van Den Bosch F. C., Maccì O A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Mundy C. J., Conselice C. J., Duncan K. J., Almaini O., Häußler B., Hartley W. G., 2017, MNRAS, 470, 3507
Murray S., Power C., Robotham A., 2013, Astron. Comput., 3-4, 23
Neistein E., Li C., Khochfar S., Weinmann S. M., Shankar F., Boylan-kolchin M., 2013, MNRAS, 15, 1
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Peng Y. J. et al., 2010, ApJ, 721, 193
Planck Collaboration XIII et al., 2016, A&A, 594, A13
Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophys. Cosmol., 4, 2
Press W. H., Schechter P., 1974, ApJ, 187, 425
Refregier A., Amara A., Kitching T. D., Rassat A., Scaramella R., Weller J., Euclid Imaging Consortium F., 2010, preprint (arXiv:1001.0061)
Salim S. et al., 2007, ApJS, 173, 267
Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14
Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, MNRAS, 128, 109
Shankar F. et al., 2014a, MNRAS, 439, 3189
Shankar F. et al., 2014b, ApJ, 797, L27
Shankar F. et al., 2015, ApJ, 802, 73
Shankar F. et al., 2017a, MNRAS, 466, 4029
Shankar F. et al., 2017b, ApJ, 840, 34
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Tinker J. L. et al., 2017, ApJ, 839, 121
Tomczak A. R. et al., 2016, ApJ, 817, 118
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
Villforth C. et al., 2017, MNRAS, 466, 812
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Wang T., et al., 2016, ApJ, 828, 56
Wetzel A. R., Tinker J. L., Conroy C., Van Den Bosch F. C., 2013, MNRAS, 432, 336 (W13)
Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., Chu Y., 2004, MNRAS, 350, 1153
Yang X., Mo H. J., Van den Bosch F. C., 2003, MNRAS, 339, 1057
Yang X., Mo H. J., Van Den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752
Zavala J., Avila-Reese V., Firmani C., Boylan-Kolchin M., 2012, MNRAS, 427, 1503
Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

APPENDIX A: ABUNDANCE MATCHING

In Fig. A1, we show the central and SSMFs from the evolving abundance matching relation described in Section 3.2. We show nine different redshifts, as labelled, using the frozen model with a dynamical-time factor $f_{\rm tdyn} = 1.0$, which best matches the local data from SDSS. The $z > 0.3$ data are from Davidzon et al. (2017).

A limitation of the present methodology is that it lacks, at present, robust observational estimates of the high-redshift central and SSMFs to better anchor our fitting procedure. Nevertheless, we see in Figs 9 and 8, the satellite SMF, the number density distribution or the fractional distribution do not necessarily require any redshift evolution in the input parameters of the SMHM relation to fit the local data. This is mainly due to the fact that most (> 60 per cent) of the accretion of the satellites observed at $z = 0$ occurs at relatively recent times, $z < 0.5 - 1$. Redshift evolution is however required for a good fit to the data when $z > 1$, as demonstrated by the grey lines in Fig. A1.

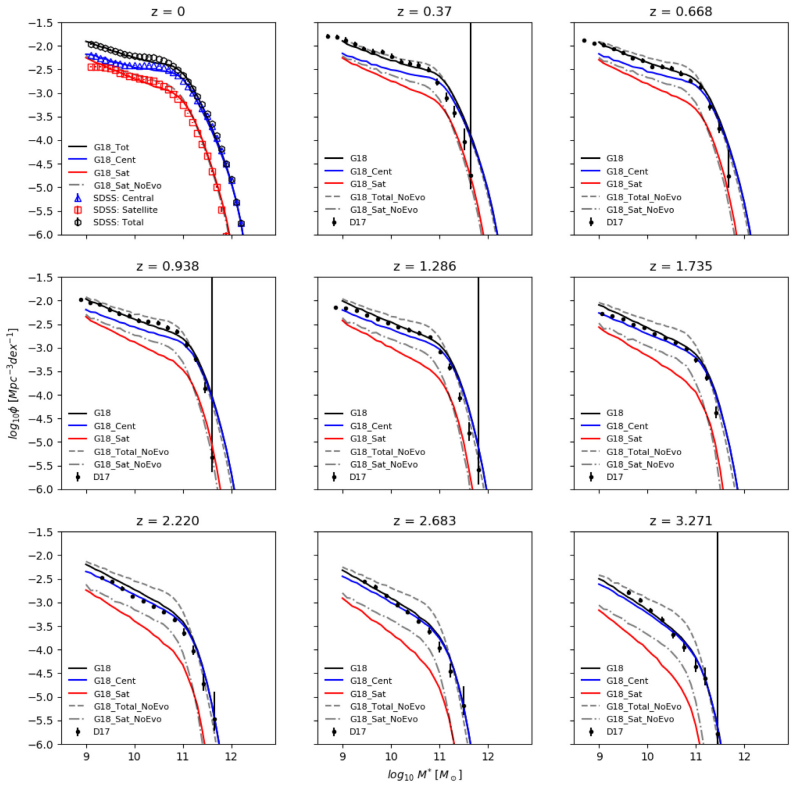


Figure A1. We show the results of the redshift dependent abundance matching at eight different epochs above redshift $z = 0$. The SDSS data are shown in the redshift $z = 0.1$ panel with crosses, higher redshift panels show COSMOS2015 (D17) data as green dots. The final result is shown in each panel as green blue and red lines for total, central, and satellite galaxies, respectively. The no-evolution is shown in grey in all panels. In this plot, G18 refers to the central SMHM fit from this work.

APPENDIX B: SFR FROM A CONTINUITY EQUATION APPROACH

A general SFR continuity equation model considers the change in the SMF over time step δt. Assuming galaxies maintain rank order the model calculates the average SFR a galaxy would have to maintain the implied growth in stellar mass. In this work, using a continuity method, as described above, we generate a SFR that is self-consistent with the HMF and abundance matching used in this work. Usually, the input SMF used is an observed quantity. However, we use as input for our continuity equation the central SMF generated by the SMHM relation from Section 3.2 and the central HMF from Tinker et al. (2010).[14] When considering the mass growth we must consider mass-loss, else the SFR is massively underpredicted. We use an instantaneous loss fraction of 40 per cent. We also neglect the mass gained from mergers. Mergers would decrease the SFR calculated (and therefore provide a better fit though out this work) so SFR from this method should be considered as an upper limit. We show in Section 3.3 our SFR has limited effect on the results of this work. We alter the parameters in equation (6) to fit our continuity equation-driven M_*-SFR relations and the new fit is given in equation (7). During this fit, we ignore the SFR inferred from galaxies that sit above the knee of the SMHM relation where the growth becomes merger dominated (Tomczak et al. 2016). We show the difference between the continuity equation approach and the Tomczak SFR in Fig. B1.

[14]We also note that given the release of resolved centrals at high redshift, improved abundance matching, and/or improved local SMF the SFR (as well as the rest of the model) is simply adjusted self-consistently as is the power of a semi-empirical model.

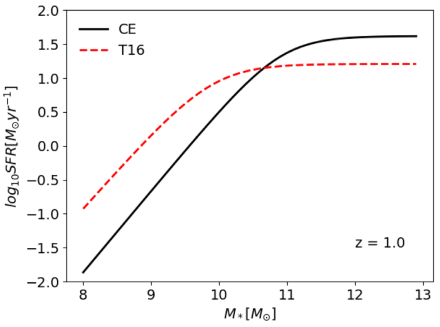


Figure B1. The Tomczak (red dashed) and continuity equation approach (black solid) SFR–stellar mass relations as given in Section 3.3.1 at redshift $z = 1$.

APPENDIX C: GLOSSARY OF TECHNICAL ACRONYMS

Throughout this work, we use several technical acronyms. We provide the reader with a glossary of these in Table C1 to aid reading of the paper and equations therein.

Table C1. A list of technical acronyms used throughout this paper.

Acronym	Definition
STEEL	The name of our model: STatistical sEmi-Empirical modeL
HMF	Halo mass function
SHMF	Subhalo mass function
USHMF	Unevolved subhalo mass function
USSHMF	Unevolved surviving subhalo mass function
δUSHMF	Unevolved subhalo mass function accretion
SMF	Stellar mass function
SSMF	Satellite stellar mass function
SFR	Star formation rate
SFH	Star formation history
MLR	Mass-loss rate

This paper has been typeset from a TEX/LATEX file prepared by the author.

MNRAS 000, 1–13 (2018) Preprint 20 January 2020 Compiled using MNRAS LATEX style file v3.0

The significant effects of stellar mass estimation on galaxy pair fractions.

Philip J. Grylls,[1]★ F. Shankar,[1]† C. Conselice,[2]‡
[1]Department for Physics and Astronomy, University of Southampton, Highfield SO171BJ, UK
[2]University of Nottingham, School of Physics and Astronomy, NG7 2RD, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

There exist discrepancies in measurements of the number and evolution of galaxy pairs. The pair fraction appears to be sensitive to both the criteria used to select pair fraction and the methods used to analyze survey data. This paper explores the connection between stellar mass estimation and the pair fraction of galaxies making use of STEEL, the Statistical sEmi-Emprical modeL. Previous results have found the pair fraction is sensitive to choices made when selecting what qualifies as a pair, for example luminosity or stellar mass selections. We find that different estimations of stellar mass such as photometric choice mass-to-light ratio or IMF that effect the stellar mass function also significantly affect the derived galaxy pair fraction. By making use of the galaxy halo connection we investigate these systematic affects on the pair fraction. We constrain the galaxy halo connection using the stellar-mass-halo-mass relationship for two observed stellar mass functions, and the Illustris TNG stellar mass function. Furthermore, we also create a suite of toy models where the stellar-mass-halo-mass relationship is manually changed. For each stellar-mass-halo-mass relation the pair fraction, and its evolution, are generated. We find that enhancements to the number density of high mass galaxies cause steepening of the stellar-mass-halo mass relation, resulting in a reduction of the pair fraction. We argue this is a considerable cause of bias that must be accounted for when comparing pair fractions.

Key words: Galaxy: halo – galaxies: abundances – galaxies: interactions – galaxies: luminosity function, mass function – galaxies: photometry

1 INTRODUCTION

ΛCDM cosmology predicts the hierarchical assembly of dark matter haloes. Throughout the history of the Universe haloes have grown in mass and size via two pathways. Firstly haloes grow via smooth accretion gradually accreting dark matter from the surrounding environment. The secondary growth mechanism is via the accretion and gradual absorption of smaller haloes, known as subhaloes. After accretion subhaloes survive as substructure of the central/host halo gradually losing mass and sinking to the center of the potential well though dynamical friction. At early epochs the gas directly follows the collapse of the dark matter sinking to the bottom of the potential wells cooling and forming stars (Mo et al. 1998). Larger dark matter haloes had a stronger gravitational influence creating a deeper potential well capturing more gas, and are thus expected to be associated, on average, with larger galaxies. The efficiency with which baryons form galaxies in haloes has been shown to depend on the halo mass, with smaller galaxies growing more with increasing halo mass than larger galaxies. The latter effect is seen in the shape of the stellar-mass-halo-mass (hereafter, SMHM) relation, where at a particular stellar/halo mass the slope of the relationship changes (e.g. Shankar et al. 2006; Moster et al. 2010; Behroozi et al. 2013; Moster et al. 2018; Shankar et al. 2014a; Conselice et al. 2018; Grylls et al. 2019b,a).

Frequent or massive mergers are thought to induce morphological changes in galaxies. Galaxies, after experiencing a massive merger, where the minor galaxy is at least a quarter of the mass of the central galaxy, are thought to lose their disk-like morphology and transform into elliptical galaxies (Negroponte & White 1983; De Lucia et al. 2006). For this reason it is important to understand the frequency and nature of mergers between galaxies to achieve a complete and coherent picture of galaxy formation and evolution. However, galaxy mergers occur on gigayear timescales and therefore it is not possible to directly observe the rate or consequence of galaxy mergers. However, by inferring the merger

★ E-mail: P.Grylls@soton.ac.uk
† E-mail: F.Shankar@soton.ac.uk
‡ E-mail: conselice@gmail.com

rate over longer timescales it can be used as a powerful constraint on cosmological parameters (Conselice et al. 2014) The traditional approach to estimate a measure of galaxy mergers is to instead count galaxy pairs at a given separation, and then assign a merging timescale to infer the rate of galaxy merging (Conselice et al. 2003, 2008; Mundy et al. 2017; Duncan et al. 2019). However, the approach of counting pairs is complicated by systematic differences when selecting galaxies, for example the evolution of the pair fraction appears to change if a selection is made by flux ratio or made by stellar mass ratio (Man et al. 2016).

In Grylls et al. (2019b) (Hereafter Paper I) we introduced a the Statistical sEmi-Empirical modeL STEEL. STEEL and other transparent semi-empirical approaches, have had multiple successes:

- Mapping galaxies into dark matter haloes using the SMHM relation, derived from the relative abundance of dark matter haloes and galaxies.
- Following the mergers of the underlying host dark matter halos.
- Computing the implied rate of galaxy mergers.

Uniquely in STEEL these properties do not follow discrete dark matter merger histories, 'merger trees', instead they are calculated using a statistical dark matter accretion history giving access to unbiased population averages. In Paper I we use this to calculate the distributions of satellite galaxies in massive haloes and in Grylls et al. (2019a) (Hereafter Paper II) this is extended to high redshift. In Paper II a comparison is made with observations of satellite distributions and stellar mass functions over multiple epochs, such we are able to confirm that the satellite distributions generated by STEEL are consistent with observations over a large range of redshifts. As STEEL is consistent with satellite distributions over multiple epochs by extension the satellite accretion histories are a reliable estimate of the true accretion.

The ability to reliably predict the galaxy merger rate is particularly powerful as it provides an additional tool to test the ΛCDM cosmological model. The galaxy merger rate is intrinsically connected to the dark matter assembly so that galaxies can be used as a proxy for the dark matter structure. However, there are notable systemics that could affect the reliability of using the galaxy merger rate from models. The primary tool in this analysis is the SMHM relation which is heavily dependent on the shape of the input stellar mass function, the comoving number density of galaxies of a given stellar mass. The stellar mass function is affected by several observational systematics, notably, choice of mass-to-light ratios, stellar initial mass functions, light profiles e.t.c... In the last decade it has been shown that the stellar mass function is significantly higher in the high mass end than previously thought, this is due to a number of improvements in stellar mass calculation including, complete fitting of the extended light profile(Bernardi et al. 2016, 2017a)), Sèrsic + exponential fitting models (Meert et al. 2015) and improved sky subtractions. Using better stellar mass estimates the number density of high mass galaxies is increased, this in turn steepens the high mass slope of the SMHM relationship (Shankar et al. 2014b; Kravtsov et al. 2018).

The effect of the stellar mass functions and the systematics introduced into the SMHM relation were investigated in Paper II. It was shown that shallow SMHM relations, associated with the previous stellar mass estimations, coupled to hierarchical ΛCDM cosmology create satellite accretion histories that are incompatible with central galaxy growth. In contrast stellar mass functions with enhanced high mass number density predict central galaxies grow more rapidly with cosmic time, such higher growth rates were found to be consistent with the total satellite accretion predicted by ΛCDM cosmology. Given the pair fraction is used as an observational estimate of satellite accretion through galaxy mergers, one would expect that where stellar mass estimations impact the accretion rates they should also produce a difference in modeled pair fractions.

In this work we investigate how the observed pair fraction changes systematically with varying SMHM relationship. We use STEEL which is capable of producing state-of-the-art satellite abundance mocks using a statistical dark matter accretion history and the galaxy-halo connection. We show that a well designed and flexible Semi-Empirical model should be used as an essential analytic tool for understanding how observational modelling assumptions, such as the estimation of stellar mass, may propagate in unpredictable ways.

This paper is laid out as follows. In Section 2 we describe the comparative simulation data used. In Section 3 we summarize STEEL the STastical sEmi-Empirical modeL and the extensions added for the analysis in this work. In Section 4 we show a systematic analysis of how the SMHM relation effects the pair fraction then compare the output of STEEL using altered SMHM relations to match simulation and observational results to show the magnitude of the differences. In Sections 5 & 6 we situate our results in a wider context and conclude.

2 DATA

2.1 Stellar Mass Functions

The input SMHM relations used in STEEL are constrained using observed stellar mass functions. STEEL uses two stellar mass functions at redshift $z = 0.1$, each using the Sloan Digital Sky Survey Data Release 7 (Abazajian et al. 2009; Meert et al. 2015), these are made using a Sèrsic-Exponential fit 'PyMorph' (Meert et al. 2015; Bernardi et al. 2016) and a de Vaucoulers fit 'cmodel' (de Vaucouleurs 1948). These stellar mass functions are representative of the previous photomotries (cmodel) and the enhanced high mass end (PyMorph) described in Section 1. To constrain the SMHM relations at higher redshift ($0.3 < z < 3.3$) we use the stellar mass functions from COSMOS2015 catalogue (Davidzon et al. 2017). These stellar mass functions are created using Bruzual & Charlot (2003) stellar population synthesis models. These stellar mass functions can be used directly with the cmodel fits, however a +0.15 dex correction is required to compare to the PyMorph light profile fitting (Mendel et al. 2014; Bernardi et al. 2013). This correction brings the redshift $z = 0.37$ stellar mass function in agreement with the PyMorph $z = 0.1$ stellar mass function matching the finding that the stellar mass function does not evolve significantly up to redshift $z = 0.5$ (Bernardi et al. 2016). The differences in the SMHM relations from these fits are discussed in Section 3.1 and fitting the parameters given in Section 3.2.

2.2 Illustris

We use data extracted from the Illustrius TNG simulation JupyterLab public data mirror (Springel et al. 2018; Nelson et al. 2018). Illustrius TNG is a state of the art, large volume, cosmological, gravo-magnetohydrodynamical simulation. To make comparisons to the results from STEEL we utilize the group and subhalo galaxy/dark matter catalogues to explore the distributions of galaxy pairs in group and cluster environments from Illustrius TNG.

3 METHOD

In this paper we highlight difficulties present when comparing models to observations, specifically pair fractions. Foremost, the difficulties stem from the assumptions inherent in a given stellar mass estimation. For example, using the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7), two notably different stellar mass functions have been derived. Using the cmodel fit (Abazajian et al. 2009) produces a stellar mass function with a sharp high mass cutoff, whereas, using a Sérsic + exponential model (Meert et al. 2015; Bernardi et al. 2016, 2017b) produces a less sharp cutoff and more high mass galaxies. The propagation of these differences into the SMHM relation, in conjunction with hierarchical assembly predicted by ΛCDM cosmology, creates different galaxy assembly histories within the same cosmology. An empirical model, such as STEEL, is ideally suited to understand these effects.

By design, an empirical model recreates the stellar mass function at all redshifts, something other modeling techniques have historically struggled with (Asquith et al. 2018). The reproduction of the stellar mass function at all redshifts is essential to derive reliable assembly histories. For example, a stellar mass function at high redshift with an excess of low mass galaxies must undergo a prolonged phase of overmerging to match the stellar mass function at low redshift. Additionally, STEEL has an advantage over more traditional models. Using a statistical dark matter accretion history, we replace discrete dark matter volumes to simulate without volume or resolution constraints. In Paper I and Paper II we showed that using the Sérsic + exponential fits, STEEL recreates the distributions of satellite galaxies over a large range of cosmic epochs. In Paper II we showed the in-situ vs ex-situ growth ratios for central galaxies for both stellar mass functions. We found that flatter SMHM relations create a satellite galaxy accretion history that is too fast with respect to the predicted growth of the central galaxy. As observations of galaxy pair fractions are often used to create an estimation of the galaxy merger rate (Mundy et al. 2017; Mantha et al. 2018). It follows that the discrepancies found in the satellite mass accretion introduced by the SMHM relation should also alter the galaxy pair fraction.

In this Paper we discuss the updates to STEEL that enable the extraction of a galaxy pair fraction given a choice of underlying stellar mass function. Using a toy model we perturb in isolation each parameter of the SMHM relation described in Section 3.2. We show that shape of the SMHM relationship directly affects the normalisation and evolution of the expected galaxy pair fraction.

In this section we begin by describing how the SMHM relation propagates into the pair fraction with the aid of two cartoons in Section 3.1. We provide the quantitative fits for the SMHM relations used in this work in Section 3.2. We then describe the statistical accretion history essential to STEEL in Section 3.3. Finally, we describe the method used to apply a spacial distribution to the statistical satellite population in Section 3.4.

3.1 Influence of the SMHM relation

The photometric choice, or other systematics such as mass-to-light ratio, the initial mass function, or background subtraction, used to calculate galaxy stellar mass, all have a direct impact on the resulting shape of the stellar mass function. For a fixed cosmological model, i.e. the number and distribution of haloes are invariant, changes in the stellar mass function directly correlate to changes in the SMHM relation. The SMHM relation can therefore be used as an effective proxy for the shape of the underlying stellar mass function.

In this work we show that different SMHM relations generate distinct pair fractions and merger rates. Stellar mass functions with greater number densities of high-mass galaxies, naturally map larger galaxies into smaller haloes due to their higher relative abundances, resulting in steeper high-mass slopes for the SMHM relations. In Figure 1 we show an illustrative cartoon of how different SMHM relations affect the galaxy mass ratios. For two identical mass halo pairs we see that a SMHM relation with a steeper slope causes a substantial difference in the stellar mass ratio mapped into the halo pairs. In general, a shallower relation causes many more massive pairs than a steeper relation.

In Figure 2 a cartoon is shown to give an example of the expected difference in the pair fraction when changing the SMHM relation. We define pair fraction as the fraction of galaxies of a given mass that have a companion with a mass equal to or greater than a quarter of the primaries mass within 5-30 kpc. The left hand column shows the SMHM relations and the right column the implied pair fractions and their evolution with redshift. In the top row we compare a steep high-mass slope to a flatter slope, where the slope has been changed at redshift $z = 0.1$. In the bottom row we compare an evolving and non-evolving high-mass slope.

The steepening of the high-mass slope increases the number of pairs created and the normalisation of the pair fraction increases. In the bottom row we show the effects of having a slope that flattens at higher redshift. We show the redshift $z = 0.1$ relation in grey and the relations with the unchanged and changed slopes in blue and orange respectively. The main effect of varying the evolution of the high-mass slope in the SMHM relation is to change the behaviour of pair fraction with redshift. A steeper slope tends to turn over the pair fraction and vice-versa. The behaviours reported in Figure 2 are what one would expect given Figure 1, where shallower slopes give higher fractions. Furthermore, from Figure 2 (and from Figure 3, shown in Section 4), it can be concluded that almost any pair fraction time dependence could be produced by appropriately altering the input SMHM relation. It is relevant to stress here that relatively minor changes in the stellar mass function can cause qualitative differences in the SMHM relation and, by extension, in

4

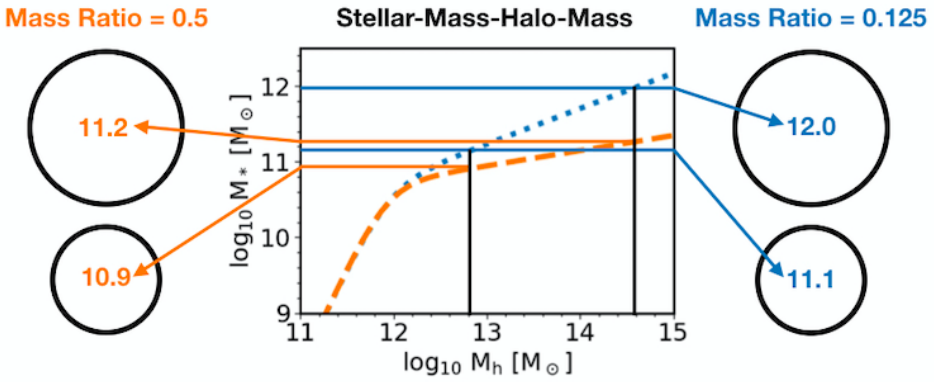


Figure 1. A diagram showing the how the SMHM relation can impact the stellar mass ratio of galaxies mapped into identical halos. The black vertical lines represent the (fixed) halo masses that are seeded with stellar mass. Coloured horizontal lines are drawn at the intersection of the halo masses and the SMHM relation these show the stellar mass seeded by each relation also shown by the numbers in matching colour on either side of the plot. The steeper SMHM relation (blue) creates a smaller stellar mass ratio as the change in halo mass maps to a much larger stellar mass difference.

the shape and normalization of pair fractions at any cosmic epoch.

3.2 SMHM relation

A double power-law relation similar to Moster et al. (2010) is used to parameterise the SMHM relation. The parameters M, N, β, and γ control, respectively, the position of the knee, the normalization, the low-mass, and the high-mass slope at redshift $z = 0.1$. Each parameter has an associated redshift evolution factor given by:

$$
\begin{aligned}
M_*(M_h, z) &= 2M_h N(z)\left[\left(\frac{M_h}{M_n(z)}\right)^{-\beta(z)} + \left(\frac{M_h}{M_n(z)}\right)^{\gamma(z)}\right]^{-1} \\
N(z) &= N_{0.1} + N_z\left(\frac{z-0.1}{z+1}\right) \\
M_n(z) &= M_{n,0.1} + M_{n,z}\left(\frac{z-0.1}{z+1}\right) \\
\beta(z) &= \beta_{0.1} + \beta_z\left(\frac{z-0.1}{z+1}\right) \\
\gamma(z) &= \gamma_{0.1} + \gamma_z\left(\frac{z-0.1}{z+1}\right).
\end{aligned}
\tag{1}
$$

We use the abundace matching fits from Paper II, shown in Table 1. The parameters for Equation 1 are given for a Sérsic-Exponential fit stellar mass function (PyMorph) and a de Vaucoulers fit stellar mass function (cmodel). The differences between these two fits are given in Section 2.1. In brief, PyMorph gives more high mass galaxies and results in a steeper high mass slope of the SMHM relation. Conversely cmodel has a lower number of high mass galaxies and creates a flatter high mass slope. For completeness, we also include a SMHM relation that well fits the outputs of the Illustris TNG which is steeper than PyMorph.

3.3 Statistical Accretion Histories

Traditional simulations such as Hydrodynamical, Semi-Analytic, or Semi-Empirical models, simulate the dark matter background of the universe using a cosmological box or a discrete set of merger trees. Both cosmological boxes and merger trees simulate a discrete cosmological volume, and in any given volume there will be a limited number of massive haloes. Due to the significant decrease in number density with increasing halo mass, haloes of even a couple of orders of magnitude smaller than the most massive halo in the simulation are found significantly more frequently. Due to the large difference in number density mergers between haloes of similar mass are extremely rare, especially compared to halo mergers with haloes with a low mass ratio. STEEL removes the dependence on discrete halo sets by using a 'statistical accretion history': haloes and mergers of any mass and mass ratio are simulated equally regardless of number density.

The full method for creating the 'statistical accretion history' is given in Paper I and a summary of the method is provided in Paper II. In brief, we start from the average growth history of central haloes and at each epoch compute the unevolved subhalo mass function[1]. At each time step the growth of the unevolved subhalo mass function is attributed to accretion of new subhaloes to create the accreted subhalo mass function (simply the number density of a given subhalo mass accreted at a given epoch for a given central mass history). Each bin of the accreted subhalo mass function is assigned a dynamical time, in essence the time to merge with

[1] This is the total number density of subhaloes accreted by a central halo over its growth history as a function of the mass ratio M_{sat}/M_{cen} where the satellite masses are frozen at infall.

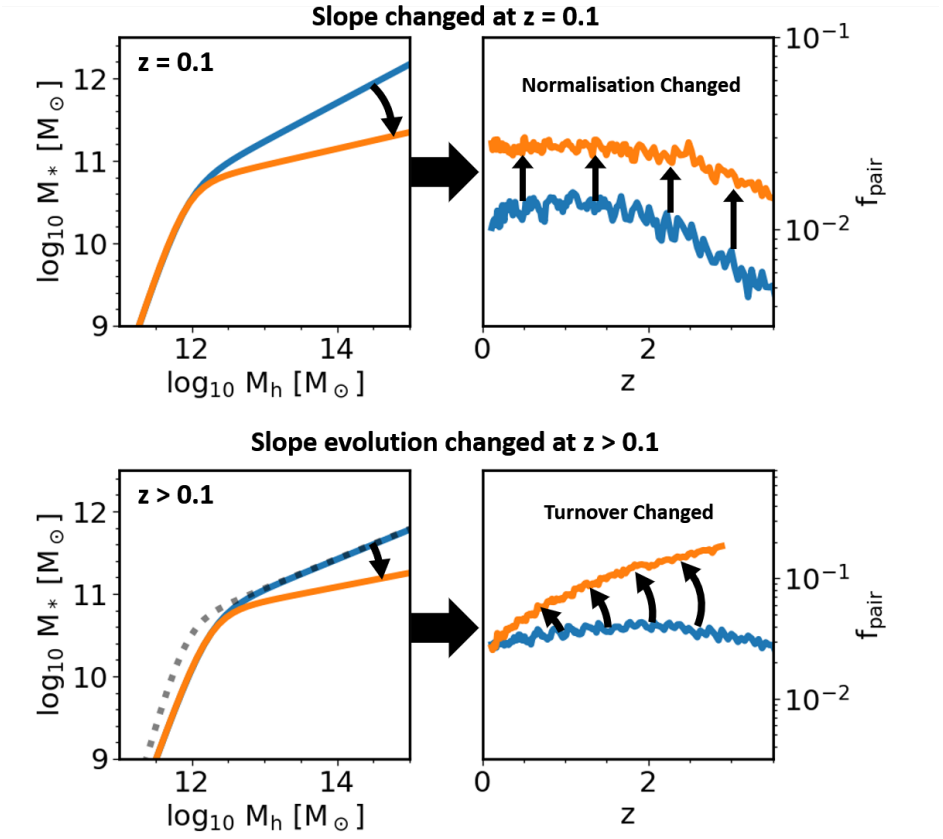


Figure 2. A diagram showing the how the SMHM relation can impact the pair fraction. The top row shows how reducing the high mass slope of the SMHM relation increases the number of pairs at all redshifts. The bottom row shows the redshift $z = 0$ relation as a grey dotted line, two relations at redshift where the relation is not evolved or evolved to be shallower are shown in blue and orange respectively. For this evolving SMHM relation the pair fractions are found to increase. In each case the reason for the increase can be explained by referencing Figure 1 where making the relation shallower seeds more massive pairs.

	M_n	N	β	γ	$M_{n,z}$	N_z	β_z	γ_z
cmodel	$11.91^{+0.40}_{-0.34}$	$0.029^{+0.018}_{-0.013}$	$2.09^{+1.21}_{-1.02}$	$0.64^{+0.11}_{-0.10}$	$0.52^{+0.24}_{-0.19}$	$-0.018^{+0.005}_{-0.004}$	$-1.03^{+0.049}_{-0.34}$	$0.084^{+0.20}_{-0.14}$
PyMorph	$11.92^{+0.39}_{-0.36}$	$0.032^{+0.016}_{-0.012}$	$1.64^{+0.85}_{-0.73}$	$0.53^{+0.11}_{-0.11}$	$0.58^{+0.15}_{0.19}$	$-0.014^{+0.007}_{-0.006}$	$-0.69^{+0.29}_{-0.36}$	$0.03^{+0.154}_{-0.147}$
TNG	11.8	0.018	1.5	0.31	0.0	−0.01	0	−0.12

Table 1. The SMHM relation fits for cmodel, PyMorph, and TNG. For the cmodel and PyMorph data, the errors are the 16th and 86th percentile from the MCMC fiting.

the central galaxies. The surviving subhalo mass function[2] is created by summing each mass of accreted subhaloes at each time step that have not yet merged with their centrals by the redshift of observation. In this work we use the PyMorph abundance matching fits from Paper II to associate each bin of the accreted subhalo mass function with a distribution of galaxies at infall. The satellite galaxies associated to subhalo bins that have reached the end of their dynamical times are summed to contribute to the average galaxy accretion history for a given dark matter halo.

3.4 Galaxy Separations

Calculation of the pair fraction requires an estimate of the distance between the central galaxy and the satellite galaxy, as we rely on our statistical accretion history, and do not have discrete halos, we assign each subhalo bin an average distance to the central galaxy. The subhaloes start at the viral radius of the central halo. The distance to the centre then reduces proportionally to the amount of dynamical time remaining (Guo et al. 2011). Throughout this work unless otherwise stated the pair fraction is calculated as the number density of satellites with a mass ratio of above 1/4 within 5 to 30 kpc of the central galaxy divided by the total number of central galaxies within the mass selection.

4 RESULTS

In this section we test the impact of two different photometric choices, PyMorph (Meert et al. 2015) and cmodel (Abazajian et al. 2009), on the pair fractions. Fixing the dark matter halo assembly and cosmology, the two photometries generate two distinct SMHM relations where the main difference lies in the high mass slope. To gain deeper insight into how the input SMHM relation propagates into the pair fractions, in Section 4.1 we show a toy model where each parameter of the SMHM relation is perturbed independently. Analysis of this toy model will inform which SMHM relation parameters affect aspects of the pair fraction, applying this in Section 4.2 we then attempt to recreate pair fractions found in the Illustris TNG simulation and in data from Mundy et al. (2017) by manipulating the SMHM relationship. Finally, using the SMHM relation that best fits the observed pair fraction we compute satellite accretion histories following the methods in Paper II. In brief, the amount of stellar mass growth expected from satellite accretion is compared to the central stellar mass growth predicted by abundance matching. If the satellite accretion is greater than the central growth the SMHM relation is non-physical and the stellar mass estimation used for the pair fraction must be questioned.

In this section we investigate the distribution of satellites around central galaxies when using different SMHM relations. This distribution is primarily dominated by the halo substructure, for this reason it is essential to make sure our selection criteria for galaxies always returns the same halo population. As we actively change the stellar masses mapped into any given halo mass one cannot use a stellar mass cut to achieve this result. From a simulation where the haloes are known, one could simply select by halo mass, however, to better match observation where haloes are not known we make a constant number density selection. A constant number density selection will always return objects that share a given number density and are therefore associated to the same haloes as we are taking the halo structure as a fixed quantity. A similar technique was employed by, Leja et al. (e.g. 2013); Mundy et al. (e.g. 2015) to trace the evolution of galaxy populations; galaxies at high redshift with a given comoving number density are assumed to be the progenitors of later populations with the same abundance. In this work we use a central stellar mass selection from the PyMorph stellar mass estimation of $M_* = 10^{11} - 10^{11.6} M_\odot$ (or $10^{9.5} - 10^{10.1} M_\odot$ when considering the low-mass slope controlled by β). The number-density range of the aforementioned mass cut is then computed and galaxies are selected from the other SMHM relations which share this number density. An example of this selection can be seen in Figure 3, the shaded horizontal band shows the stellar masses for each SMHM relation that share number density.

4.1 Systematic Analysis

We use the flexible nature of STEEL to create a toy model where each of the main parameters (M, N, β, γ), and their evolutionary factors (M_z, N_z, β_z, γ_z), governing the SMHM relation are adjusted in turn to explore the affect on the galaxy pair fractions. Table 2 details the change made to the SMHM relation for each parameter.

Figure 3 shows each of the SMHM relations in the outer four panels, the reference SMHM relation PyMorph is shown in blue at redshifts $z = 0.1$ (dotted line) and $z = 2$ (dashed line) in each panel, the modified redshift $z = 0.1$ relation is then shown in orange, and the increased and decreased (dashed red and green) evolution are shown at redshift $z = 2$. The inner four panels follow the same colour convention. When changing M, the knee parameter, a large increase in the pair fraction is found from a lower knee: The shallower high mass slope is extended therefore more haloes are seeded in the mass range for pairs. We see the same effect at high redshift, the lower value of M at high redshift creates a higher pair fraction.

The normalization parameter, N, creates little change in the pair fraction as expected because the mass ratios are largely unaffected. The low mass slope parameter, β, affects the seeding of smaller galaxies hence a lower mass range is used for the consistent number density cut. Due to the steepness of the low mass slope the fraction of pairs is lower in this mass cut. Finally, when the high mass slope parameter, γ, is altered more pairs are found at high and low redshift when the slope is shallow. This is again attributed to more galaxies seeded within the mass ratio range.

[2] The total number density of subhaloes that one would expect to still be present in the parent halo at a given epoch.

Table 2. The adjustments to the SMHM relation used in Figure 3.

	PyMorph	$X_{0.1,alt}$	$X_{z,+}$	$X_{z,-}$
M	11.92	-0.25	-	-
M_z	0.58	-	+0.1	-0.1
N	0.032	+0.04	-	-
N_z	-0.014	-	+0.007	-0.007
β	1.64	-0.3	-	-
β_z	-0.69	-	+0.3	-0.3
γ	0.53	+0.06	-	-
γ_z	-0.03	-	+0.2	-0.2

8

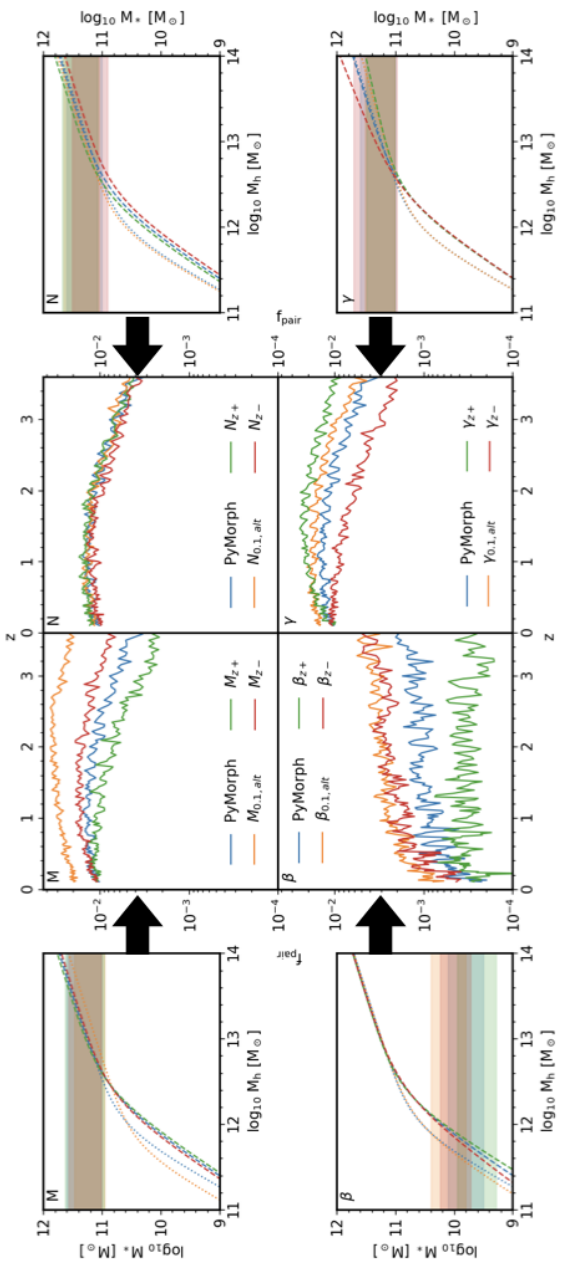


Figure 3. Each of the panel pairs (M, N, β, γ) shows the input SMHM relation in the outer plot and the modelled pair fraction evolution in the center plot. Each pair investigates adjustments to the given parameter of the SMHM relation (M, N, β, γ). Each pair shows the reference SMHM relation 'PyMorph' in blue, the relation adjusted at redshift $z = 0.1$ keeping the same SMHM relation evolution parameters in yellow. The red and green lines respectively have the evolution parameter altered such that the evolution parameter is increased or decreased with respect to the PyMorph relation from Sérsic Exponential fits (Meert et al. 2015). In the outer (SMHM relation) plots dotted lines are $z = 0.1$ relations and dashed lines are $z = 2$ relations the PyMorph is shown at both epochs for comparison. Finally the shaded bands in the outer plots show the consistent number density selections used in the center plots.

4.2 Simulation and Observational Results

The observed pair fraction is known to have discrepancies based on the galaxy property used to calculate the ratio. In Man et al. (2016) it is shown that selecting pairs by flux ratio or stellar mass creates differences in the pair fraction evolution. In Figures 1 & 2 we show the predicted effect of the determination of stellar mass on the SMHM relation and the propagation of these changes into the pair fraction. Through the use of a toy model in Figure 3 we show how isolated perturbations to the eight SMHM relation parameters propagate into the galaxy pair fraction. Given this analysis we find that any observation of the pair fraction must be understood in terms of its implicit observational assumptions. Furthermore, direct comparison of pair fraction results should only be undertaken under identical stellar mass derivation assumptions, where this is not the case the influence of any differences must be accounted for. In this section we fit, by making use of STEEL, observed pair fractions using small changes to the SMHM relation. We anticipate this modelling can be used to provide corrections to pair fraction results to allow for fair comparisons.

In Figure 4 we show the simulated galaxy pair fractions for galaxies in the mass range $M_* = 10^{10}$ to $10^{10.6}$. The pair fraction is shown for two different SMHM relation inputs to STEEL. In blue we show the PyMorph (Sèrsic Exponential) input used as the baseline in Figure 3, in orange the input calibrated to match the Illustris TNG simulation. In the right hand panel we see that the the prediction from STEEL with the TNG-calibrated input is in good agreement to the pair fraction extracted directly from the Illustris TNG simulation. The pair fraction predicted using the PyMorph input is 0.5 dex lower, this is to be expected as in the mass range we are considering the Illustris TNG simulation SMHM relation is shallower and more pairs are therefore created in a greater mass range of halo mergers.

Figure 5 shows the predicted pair fraction evolution using the two SMHM relations from PyMorph and cmodel presented in Section 3.2. The left panel shows each SMHM relation at redshift $z = 0.1$ and $z = 2.5$. Following the systematic investigation in Figure 3 we attribute the 0.1 dex difference in pair fraction to the difference in high mass slope between PyMorph and cmodel. The best-fit relation from Mundy et al. (2017), shown as black crosses, rises rather than falling as seen from PyMorph and cmodel. We see in Figure 3 a SMHM relation with a high mass slope that decreases with redshift creating a pair fraction evolution of this nature.

To attempt to match the Mundy et al. (2017) pair fractions we begin using the cmodel SMHM relation which gives the closest match in pair fraction at low redshift. Following the analysis of Figure 3 where higher γ_z increases the pair fraction at high redshift we alter the parameter from 0.0 to 0.5 in steps of 0.1. In Figure 6 the left panel shows the SMHM relation at redshift 0.1 as a black dotted line then coloured lines show the relation at redshift $z = 2$ given the different γ_z parameters. The right panel shows the impact of this evolution on the pair fraction, as predicted higher γ_z increases the pair fraction with redshift and a value of above 0.1 removes the turnover. Comparing to Mundy et al. (2017) we see a value of γ_z between 0.1 and 0.2 best reproduces the rise in pair fraction.

In Paper II we show not all SMHM relationships are internally self-consistent within a given ΛCDM assembly. Following the average growth of a halo mass bin one can use the SMHM relationship to predict the average galaxy growth for this population of haloes. The average satellite accretion should then be less than, or at most equal to, the central stellar mass growth both in total accretion and instantaneous rate. In Figure 7 we show the self-consistency in terms of stellar mass for four fits. We select galaxy populations at three stellar masses as indicated by the column labels. The galaxies are chosen at redshift $z = 0.1$ and by are followed along the average halo growth histories. Each halo growth history is mapped to an average stellar mass history using abundance matching. For each evolution history STEEL calculates the satellite accretion over the galaxy/haloes' history.

The four fits shown are, the cmodel fit, then the two γ_z models that are closest to the Mundy et al. (2017) pair fraction, and the maximum slope evolution tested $\gamma_z = 0.5$. The top row shows the total mass predicted by abundance matching (solid lines) and from satellite accretion (dashed lines), the middle row shows the ratio of mass accreted to total mass gained since redshift $z = 3$, the bottom row shows the ratio of instantaneous satellite accretion to instantaneous growth rate. In the middle and bottom rows the solid black lines are at unity and a model that goes above this line is nonphysical as more matter would have been accreted than the galaxy growth history can account for. From Paper II we know cmodel to be internally inconsistent, for $\gamma_z = 0.5$ the slow growth at redshift $z = 2$ means satellite accretion is far more rapid than galaxy growth, the $\gamma_z = 0.1$ and 0.2 models are fairly simmilar with the 0.1 model slightly favoured, the total accretion and ratio are good (top and middle rows, green line) the instantaneous rate is slightly high but could be accounted for via a greater loss of mass to the intra cluster medium during a merger[3]. The models with γ_z that give the closest fit to Mundy et al. (2017) are also closest to internal consistency. This has two implications: Firstly, an evolving high mass slope can create the central galaxy growth required for consistency with the satellite accretion predicted by ΛCDM cosmology. Secondly, it suggests that the pair fraction can provide an additional constraint on the SMHM relation, in particular we see here the evolution of the pair fraction can constrain the evolution of the high mass slope.

5 DISCUSSION

The primary goal of this paper is to show the propagation of systematics in galaxy modelling. Specifically, we use the SMHM relationship to connect assumptions used when estimating stellar masses from observations to systematics in galaxy pair fractions in the context of a ΛCDM Universe. In this work we have used two observed stellar mass functions from SDSS-DR7 observations that use a de Vaucoulers and a Sèrsic + Exponential fit to determine stellar masses, which generate stellar mass functions with notably different number densities at the high mass end. Each stellar mass

[3] The models presented here use $f_{loss} = 0.5$ similar to Moster et al. (2018).

10

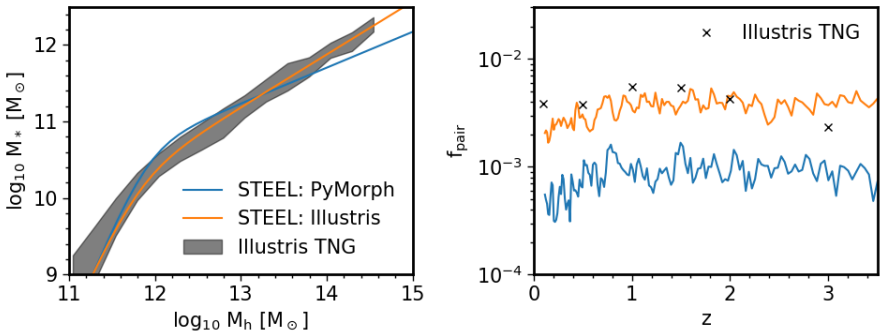


Figure 4. Left: Two SMHM relations are shown from STEEL using parameters designed to reproduce the SMHM relation found in the Illustrius TNG simulation (Orange line) and the PyMorph(Sèrsic Exponential) fit parameters (Blue line). The shaded region is the output from the Illustris TNG simulation. Right: The pair fraction, for galaxies in the mass range $M_* = 10^{10} M_\odot$ to $10^{10.6} M_\odot$ generated from STEEL is shown for runs using both the SMHM relations, lines follow the same colours as the left hand panel. The pair fraction extracted directly from the Illustris TNG simulation is shown using black crosses.

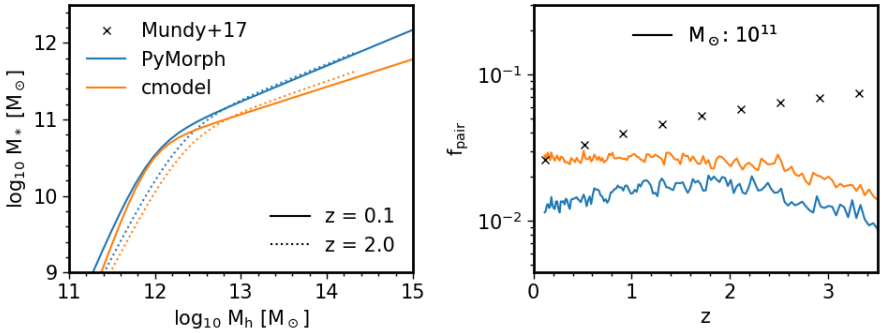


Figure 5. Left: The stellar mass halo mass relations derived from PyMorph (blue) and cmodel (orange) at redshifts 0.1 (solid lines) and 2.0 (dotted lines). Right: The pair fraction evolution for galaxies using both SMHM relations. We make mass cuts, $> 10^{10} M_\odot$ (dashed line) and $> 10^{10} M_\odot$ (solid line), in PyMorph and cmodel. The black crosses show the corresponding best fits for the $> 10^{11} M_\odot$ mass cut from Mundy et al. (2017).

function generates, through abundance matching, a different SMHM relationship. The Sèrsic + Exponential mass function generates a steeper high-mass slope in the SMHM relationship at any epoch. In addition to the SMHM relationships from the observed data we use a relationship fitted to match the outputs of the Illistris simulation. Furthermore, we also consider a toy model SMHM relation individually perturbing each input parameter to transparently probe the impact of the input SMHM relation on the pair fractions. In each case we find that small changes introduced into the SMHM relationship can have significant effects on the expected pair fractions, as shown in Figures 1,2, & 3. This suggests that in the contxet of a ΛCDM Universe tensions in previous observational studies could, in large part, be traced back to systematics in stellar mass estimates.

In Mundy et al. (2017) the $M_{*,cen} > 10^{10}$ pair fraction is given, however, this is not significantly different from the $M_{*,cen} > 10^{11}$ pair fraction shown in Figure 5. In Figure 3 we find the pair fraction drops significantly when mass a selection is taken below the SMHM relation knee. As this drop

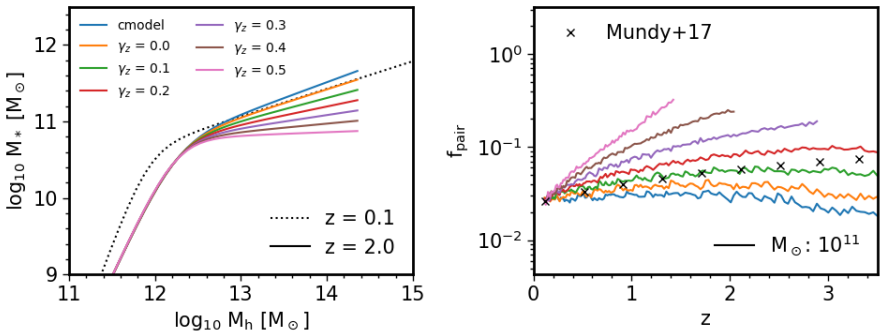


Figure 6. Left: The stellar mass halo mass relations derived from cmodel (black) at redshift $z = 0.1$ (dotted lines) and at $z = 2.0$ (coloured lines) with altered high mass slope evolution parameter. Right: The pair fraction evolution for galaxies each altered SMHM relation. The black crosses show the corresponding best fits for the $> 10^{11} M_\odot$ mass cut from Mundy et al. (2017).

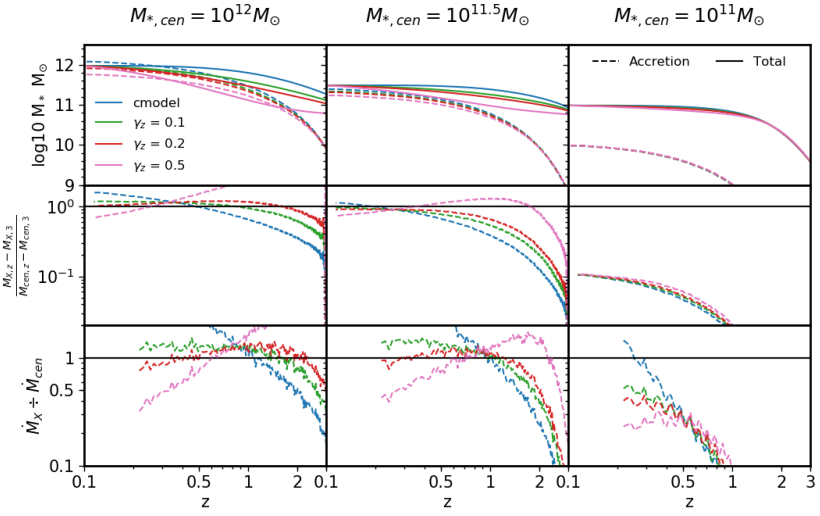


Figure 7. Average 'mass tracks' are shown which have central galaxy masses at redshift $z = 0.1$ of $M_{*,cen} = 10^{12}$, $10^{11.5}$, and 10^{11} $[M_\odot]$ from left to right. The satellite galaxy accretion is shown for evolved satellites with a dashed line. The top panels show the total mass of the central (solid lines) and the total mass gained from accretion. The middle panels show the fraction of the total galaxy mass formed from satellite accretion since redshift $z = 3$. The bottom panels show the ratio of the mass accretion rate from satellite galaxies to the mass growth rate of the central galaxy predicted by abundance matching. The black horizontal lines in the second and third rows are at unity. The colours are coded to the high mass slope evolution parameter as shown in the legend.

is not found by Mundy et al. (2017) we interpret that their pair fraction measurement is not consistent with a break in the SMHM relation between $10^{10} M_{s,cen}$ and $10^{11} M_{s,cen}$.

Man et al. (2016) noticed that the choice between luminosity-selected and stellar-mass selected pairs affected the pair fraction evolution. In this work we have provided a clear framework to properly interpret how input choices create systematic effects in the observed pair fraction and its evolution. Furthermore, it is a common approach to infer the assembly history of galaxies by converting the pair fractions into merger rates by assigning timescales to galaxy pairs (Conselice et al. 2003, 2008; Mundy et al. 2017). In Paper II we developed a model that calculates the stellar mass growth rates of central galaxies and the stellar mass accretion rate from satellite galaxies. It is found that, for some SMHM relationships, the accretion rate can be greater than the total growth rate implying the model is internally inconsistent and the SMHM relationship is not compatible with this ΛCDM cosmology. The stellar mass accretion rate is connected to the merger frequency and therefore the galaxy pair fraction. In this work we connect the shape and evolution of the SMHM relationship to the evolution of the pair fractions. We propose it is therefore possible to use the pair fraction as an additional constraint to the SMHM relationship, this is a natural extension of conditional abundance matching or extended SHAM (subhalo abundance matching) models (Hearin et al. 2013). Using STEEL one can test simultaneously the accretion ratio and the pair fraction generated from a given stellar mass function and cosmology.

Any changes to the stellar mass estimates such as photometry, background subtraction, IMF, e.t.c. that affect the stellar mass function will in a given cosmology create a change in the SMHM relationship. Therefore by the systematic propagation demonstrated in this work any stellar mass estimation will create systematic differences in the pair fractions. With the techniques presented in this work and in Paper I & Paper II, one could retrieve the systematic differences created in pair fraction under multiple ΛCDM cosmologies and for any given set of stellar mass functions. As a further test of our results we show the pair fractions found in Illustris, together with the pair fractions predicted by STEEL, using a SMHM relationship designed to match that of Illustris are fully consistent. Using STEEL the pair fractions produced from different ΛCDM cosmologies can be tested to determine which best fits the observed pair fraction when ensuring self consistency in stellar mass estimates is maintained throughout.

In the era of wide and deep surveys, such as EUCLID, constraining a model using a single multi-epoch data set with consistent photometry will become a reality. The advantages of this are twofold: By tuning the SMHM relation to a given survey over a large range of redshifts the growth of the stellar mass function over time can be tested against the implied satellite accretion and star formation rate as in Paper II this can be seen as a test of the consistency of the cosmological model or of the consistency of the stellar mass and/or starformation rate estimation. Secondly, as in Paper I one can test if the high redshift SMHM relation produces the low redshift satellite distributions. The constraints on a given photometry, cosmological model, satellite evolution, starformation rate, e.t.c... are still not complete however it will allow nonphysical results to be identified. Furthermore, by making the model accessible it can then be used in the manner described above to make systematic adjustments to compare between current and future data sets that may use different stellar mass estimations.

6 CONCLUSIONS

In this paper, we show that the input SMHM relations, based on different stellar mass estimations, have a significant impact on the predicted galaxy pair fraction. In short, the steeper the relation, the lower the pair fraction. Specifically we compare stellar mass functions created with a de Vaucoulers-based photometry (cmodel) to a Sérsic-Exponential photometry (PyMorph), the latter leading to an enhancement in the number density of high mass galaxies. The resulting effect of these stellar mass functions is a different input SMHM relation to STEEL, the primary difference consisting of a steeper high mass slope when adopting the Sérsic-Exponential profile. As expected, the Sérsic-Exponential results in a lower pair-fraction. To attempt to explain the difference in pair-fraction evolution with redshift we create a suite of toy models testing different alterations to the SMHM relation. We find that this evolution is linked to the evolution of the high mass slope.

The purpose of this paper is to show how subtle changes in derivation of stellar mass could lead to large differences in the pair fraction observed. This work is particularly important in the context of understanding galaxy mergers, the rate of which is commonly inferred from pair fractions, and the effects thereof. Merger rates are used to: predict rates of mass accretion, invoke variations in morphological type in many galaxy models, and are thought to relate to the triggering of AGN. It is therefore crucial when comparing two different samples to account for the systematic biases introduced by the assumptions implicit in stellar mass estimation as shown in this work. Future surveys should look to use fast and flexible modelling alongside data products to be able to properly understand the systematic effects of assumptions made on derived data products. For example the SMHM relation must simultaneously fit: traditional abundance matching, self-consistency between satellite accretion and central galaxy growth, and as shown in this paper the normalisation and evolution of the galaxy pair fraction. This multi-product fitting will ensure relations such as the SMHM relation are better constrained. The methods described in this work and Paper II will also provide more stringent theoretical limits to the stellar mass estimations and photometerys used for future surveys.

ACKNOWLEDGEMENTS

We acknowledge extensive use of the Python libraries astropy, matplotlib, numpy, pandas, and scipy. PJG acknowledges support from the STFC for funding this PhD. FS acknowledges partial support from a Leverhume Trust Research Fellowship.

REFERENCES

Abazajian K. N., et al., 2009, The Astrophysical Journal Supplement Series, 182, 543

Asquith R., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 1197

Behroozi P. S., Wechsler R. H., Conroy C., 2013, The Astrophysical Journal, 770, 57

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, Monthly Notices of the Royal Astronomical Society, 436, 697

Bernardi M., Meert A., Sheth R. K., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016, Monthly Notices of the Royal Astronomical Society, 455, 4122

Bernardi M., Meert A., Sheth R. K., Fischer J. L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2017a, Monthly Notices of the Royal Astronomical Society, 467, 2217

Bernardi M., Fischer J. L., Sheth R. K., Meert A., Huertas-Company M., Shankar F., Vikram V., 2017b, MNRAS, 468, 2569

Bruzual G., Charlot S., 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000

Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003, Astronomical Journal, 126, 1183

Conselice C. J., Rajgor S., Myers R., 2008, Monthly Notices of the Royal Astronomical Society, 386, 909

Conselice C. J., Bluck A. F. L., Mortlock A., Palamara D., Benson A. J., 2014, Mon. Not. R. Astron. Soc, 444, 1125

Conselice C. J., Twite J. W., Palamara D. P., Hartley W., 2018, The Astrophysical Journal, 863, 42

Davidzon I., et al., 2017, Astronomy & Astrophysics, 605, A70

De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, Monthly Notices of the Royal Astronomical Society, 366, 499

Duncan K., et al., 2019, The Astrophysical Journal, 876, 110

Grylls P. J., Shankar F., Leja J., Menci N., Moster B., Behroozi P., Zanisi L., 2019a, Monthly Notices of the Royal Astronomical Society

Grylls P. J., Shankar F., Zanisi L., Bernardi M., 2019b, MNRAS, 483, 2506

Guo Q., et al., 2011, Monthly Notices of the Royal Astronomical Society, 413, 101

Hearin A. P., Zentner A. R., Berlind A. A., Newman J. A., 2013, Technical Report 0000, SHAM Beyond Clustering: New Tests of Galaxy-Halo Abundance Matching with Galaxy Groups, `http://www.multidark.org`. `http://www.multidark.org`

Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, Astronomy Letters, 44, 8

Leja J., van Dokkum P., Franx M., 2013, The Astrophysical Journal, 766, 33

Man A. W. S., Zirm A. W., Toft S., 2016, The Astrophysical Journal, 830, 89

Mantha K. B., et al., 2018, Monthly Notices of the Royal Astronomical Society, 475, 1549

Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943

Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, The Astrophysical Journal Supplement Series, 210, 3

Mo H. J., Mao S., White S. D. M., 1998, Monthly Notices of the Royal Astronomical Society, 295, 319

Moster B. P., Somerville R. S., Maulbetsch C., Van Den Bosch F. C., Maccì O A. V., Naab T., Oser L., 2010, The Astrophysical Journal, 710, 903

Moster B. P., Naab T., White S. D. M., 2018, Monthly Notices of the Royal Astronomical Society, 477, 1822

Mundy C. J., Conselice C. J., Ownsworth J. R., 2015, Monthly Notices of the Royal Astronomical Society, 450, 3696

Mundy C. J., Conselice C. J., Duncan K. J., Almaini O., Häußler B., Hartley W. G., 2017, MNRAS, 470, 3507

Negroponte J., White S. D. M., 1983, Monthly Notices of the Royal Astronomical Society, 205, 1009

Nelson D., et al., 2018

Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, The Astrophysical Journal, 643, 14

Shankar F., et al., 2014a, Monthly Notices of the Royal Astronomical Society, 439, 3189

Shankar F., et al., 2014b, The Astrophysical Journal, 797, L27

Springel V., et al., 2018, Monthly Notices of the Royal Astronomical Society, 475, 676

de Vaucouleurs G., 1948, Annales d'Astrophysique, 11

This paper has been typeset from a TEX/LATEX file prepared by the author.

Appendix E

Fellowship application and future of STEEL

This is PJG's unsuccessful but well reviewed (4/6, 4/6, 6/6) fellowship application. Included are original submission documents. These documents represent PJG's technical vision for the future of STEEL.

Hawking Fellowship: Case for support

1 Track Record

I have achieved exceptional results at all stages of my education. At my highly competitive school I received the award for the best student, for achievement and engagement, in every subject I took at A-Level. I gained entry to the University of Bath where I studied Physics with a strong computational component gaining 1st Class Honours. I went on to do a Post Graduate Certificate of Education gaining valuable teaching and communication experience.

I began my PhD in Astrophysics in 2016 under the supervision of Dr Francesco Shankar. I began by working on a preliminary cosmological semi-empirical model, after gaining understanding of the field and modelling techniques I began to design and experiment with my own methodology. Current techniques have consistently struggled to faithfully reproduce the distributions and number densities of galaxies over multiple redshifts (Asquith et al., 2018). There are two major issues facing current models; firstly, over parameterization of models introducing moderate to severe degeneracies in the solutions (e.g. Lapi & Cavaliere, 2011). Secondly, these models work by simulating a discrete number of dark matter haloes which introduce a volume and resolution bias. The volume bias is due to the relative abundance of small to large haloes, for example to simulate ten massive haloes several thousand or even millions of smaller haloes must also be simulated. The resolution bias is due to the necessary cutoff in the smallest halo size due to the aforementioned high number density, in a large simulation the resolution must be limited to not exceed computational resource. These trade-offs mean it has been hard to simulate with good statistics of massive haloes without a low resolution cutoff to prevent simulating excessive numbers of small haloes. STEEL is a beyond the state-of-the-art technique, which I designed to better model satellite galaxies in groups and clusters by overcoming the issues suffered by traditional techniques. I list the innovations I have made with STEEL that are presented in Grylls et al. (2019) and my two upcoming papers (submitted and in prep).

- Developed a novel 'Statistical Dark Matter Accretion History' to remove volume and resolution bias from my simulations. Unconstrained by volume STEEL is able to model massive clusters missed by traditional models such as the remarkable high redshift cluster presented in Wang et al. (2016).
- Used the accretion histories and abundance matching to show to first order that dynamical time and not environmental processes effect distributions of satellite galaxies.
- Used the unique ability of a statistical semi-empirical model simulating the population averages to show that galaxy stellar mass estimates are inconstant with galaxy assembly models.

During my PhD I given talks in a variety of contexts, from a Lorentz centre workshop to the non-cosmology Euclid session at NAM 2017, I have also given an invited seminar to discuss STEEL at the University of Nottingham. I have co-supervised two summer students working on lenticular galaxies to create and implement empirical routines to model these elusive galaxies.

2 Research Proposal

Dark matter is ˜85% of the matter in the universe (Planck Collaboration et al., 2015); Leading theory states that the structure formation of dark matter is directly connected to the formation of galaxies. However, dark matter cannot be directly observed therefore understanding of the dark matter cosmological

paradigm is limited by the quality of observations that can be used as indirect measurements. It is expected that each dark matter halo forms a galaxy at its core. It is therefore possible to reveal the evolution of dark matter structure by tracing the formation and distribution galaxies across cosmic time. It is however unfortunate that major systematic errors in the analysis of galaxy observations undermine confidence in using galaxies as a tracer. The foremost systematic is in the estimation of galaxy stellar mass, I have shown using STEEL that current mass estimates are not consistent with dark matter assembly history. In this application I describe how I have highlighted and constrained this effect as well as the successes I have had in using STEEL to make self consistent predictions. This proposal shows how I can extend STEEL to ensure the self consistency of stellar mass estimates and dark matter assembly in future surveys. This work will be able to put constraints simultaneously on dark matter cosmological models, fitting models for data, and theories of galaxy assembly.

2.1 State of the art and current limitations

The leading cosmological theory that describes the formation and distribution of galaxies is the ΛCDM cosmological model (see Bull et al., 2016, for a review). Notable successful predictions of ΛCDM cosmology include the Lyman-α forest, galaxy clustering and weak gravitational lensing. The traditional problems such as the 'cusp-core', missing satellites and 'too big to fail' have been mostly addressed by better simulations (coupling baryonic feedback at high resolution), better observations (observing faint satellites) or alternative theories (self interacting dark matter).

There are no current techniques to directly measure dark matter, observational tracers such as the distribution of galaxies are therefore key to our current understanding of dark matter structure. The current state-of-the-art method is to model the co-evolution of dark matter and baryonic matter (gas and stars that make up galaxies) to create galaxy mocks. The two leading techniques are Hydrodynamic simulations and Semi-Analytic models each uses a different method to include the dark matter component. In numerical simulations it is found that dark matter has a hierarchical evolution, from small perturbations in the initial density field large haloes grow under the influence of gravity smoothly accreting more dark matter as well as accreting other haloes in merger events. Hydrodynamical galaxy simulations co-evolve baryons directly with a dark matter simulation such that the gravitational influence of the baryons, as well as other feedback effects, can influence the dark matter. Semi-Analytic models use either the dark matter component from hydrodynamical simulations or an N-body dark matter only simulation to create merger trees[1], a simplification of the complex merging structure of dark matter.

In hydrodynamic simulations using a large number of cells the system is evolved by solving the equations of hydrodynamics, any process that would fall below the resolution of a given cell is then solved using sub-grid methods which are analytic routines that can be tuned to produce results that agree with observations. Hydrodynamical simulations are powerful tools and are able to give many physical insights into galaxy formation as they directly resolve structure formation and the output galaxies. However, the foremost limitation of such modelling is computational resource, each simulation can take many months to run, it is therefore required to make compromises in either the volume or resolution of the simulation. Furthermore, the sub-grid tuning can introduce degeneracies into the model obscuring the actual physical processes.

Semi-Analytic models initialise dark matter haloes with gas at high redshift then follow analytic routines to evolve the baryonic matter. The mergers of the galaxies are predominately dictated by the dark matter histories from dark matter merger trees. A major component of a Semi-Analytic model is the number of parameters that can be tuned to reproduce observations, less computationally expensive than Hydrodynamic simulations, Semi-Analytic models can be run many times to gain the best fit. However, as with hydrodynamical simulations such tuning can lead to degeneracies that obscure the actual physics.

[1] Alternatively analytic routines have been created to 'grow' merger trees that mimic simulations that grant flexibility in volume and/or cosmological parameters without the need to run multiple large volume simulations Parkinson et al. (2008)

Semi-Empirical models are a potential solution to the degeneracy introduced by the multi-parameter tuning inherent to the aforementioned techniques. Semi-Empirical models closely link observations to theory, in this instance ΛCDM dark matter merger trees, to create a model that by design reproduces specific observations. Commonly abundance matching, a technique used to predict the amount of stellar mass expected for a given halo mass, is used to populate dark matter haloes to reproduce the observed stellar mass function over multiple epochs. Additional physical assumptions are then gradually introduced adding the least assumptions/tuning parameters required to obtain a model consistent with observations. Building a model cautiously introducing additional assumptions/parameters limits the degeneracy from over parameterization improving the transparency of the model with regard to the important physics. Volume and resolution remain limitations of traditional Semi-Empirical models as they are built on the same dark matter merger trees as Semi-Analytic models.

My model STEEL is the first STastical sEmi-Empirical modeL. It builds on the low parameter modelling that make Semi-Empirical models powerful and further introduces a *statistical dark matter accretion history* removing the limitations inherent in traditional dark matter merger trees. STEEL has proven its potential with successes including accurate reproduction of the distribution of satellites in central haloes and the prediction of galaxy star formation rate. In addition to this the unique statistical nature of STEEL has allowed for the investigation of the self consistency of galaxy accretion in a hierarchical cosmological model. Removing discrete galaxies I explore the average satellite galaxy accretion histories and average central galaxy growth histories over the entire population. I find that many traditional galaxy mass estimates produce accretion rates that are inconsistent with central galaxy growth, I am therefore able to conclude that the hierarchical evolution of the dark matter and/or the galaxy mass estimates are flawed.

I propose STEEL be updated to become part of data processing pipelines to check derived products from observations in a given cosmology to ensure fitting model and cosmological model consistency. Once implemented this will be a **fundamental redesign to the connection between theoretical models and observational fitting in extra-galactic astrophysics**. Given the multitude of missions set to begin in the near future it is essential that models we use to fit data and theoretical models of galaxy formation are in agreement. Disagreement in these models is either a failure in the theory, observational fitting methods, and/or, an indication that ΛCDM assembly is incorrect. Results from STEEL will confirm or reject cosmology/fitting for surveys such as Euclid, an extra galactic mission with the main goal of providing precise cosmological constraints.

2.2 Motivation

The outcomes of this project are twofold: Firstly, STEEL will continue to be developed as a galaxy evolution model that will make testable predictions about galaxy populations derived from the dark matter accretion histories. Secondly, STEEL will be expanded in capability to be able to actively test the consistency of observational fitting models and ΛCDM cosmology. Through each of these pathways I will be able to place constraints on the ΛCDM cosmological model. Using STEEL I will also be able to test and develop theories of galactic formation and satellite evolution, with particular emphasis on the connection between mergers, active galactic nuclei, and quenching (the reduction of star formation rate). Finally, by using STEEL to investigate the systematics introduced by different cosmologies and observational fitting models I will be able to re-normalise data products from different surveys to correct for these effects.

2.2.1 Methodology of STEEL

STEEL can be broadly divided into four distinct aspects: Observational Data, Cosmology/Dark Matter, Galaxy Modelling, and, Outputs/Predictions. Figure 1 shows a cartoon of how steel works subdivided in these four categories. STEEL has successfully investigated satellite distributions, pair fractions and, star formation rates, however, for this cartoon we focus on the star formation rate as it best shows the

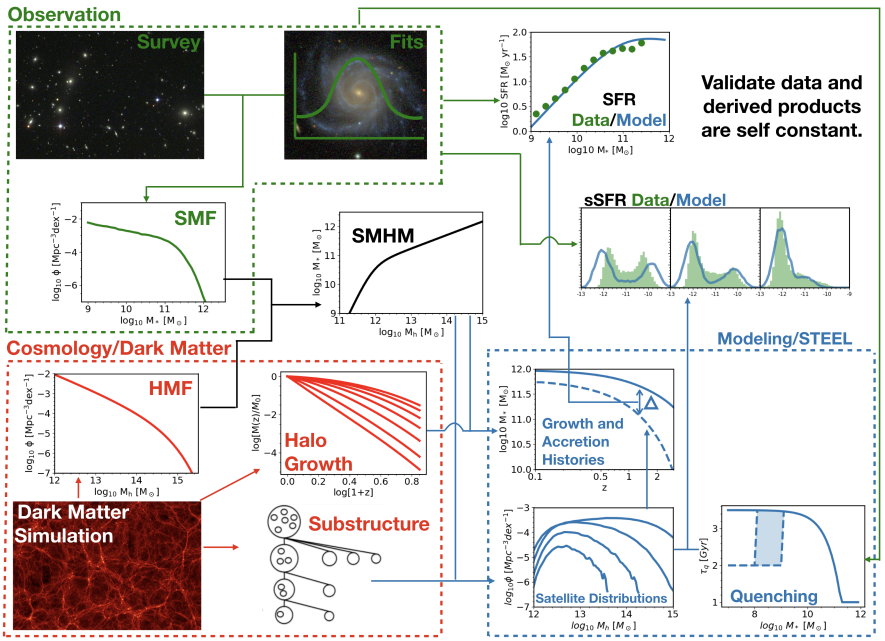


Figure 1: Schematic cartoon of how STEEL empirically models star formation rates. Described in full in Section 2.2.1.

interconnections of the empirical model. (In the following text colour directly corresponds to sections of the Figure 1.) By fitting and counting galaxies from surveys one can create the stellar mass function (SMF), the number density of galaxies as a function of mass. Using a halo finder to identify dark matter haloes a halo mass function (HMF), the number density of dark matter haloes as a function of mass, is extracted from dark matter simulations. Galaxies are assumed to reside in haloes with the same number density implying a mapping between galaxy stellar mass and host halo mass, this relation named the stellar-mass-halo-mass relation (SMHM) is the key empirical input used in STEEL . Using the stellar-mass-halo-mass relation and the dark matter substructure it is possible to derive the satellite distribution over multiple epochs. Using theoretical dynamical time arguments the satellite accretion history can be calculated, in addition the the central galaxy growth history is calculated using the stellar-mass-halo-mass relation and the halo growth histories. The difference between the the central galaxy growth history and the satellite accretion history can be attributed to galaxy growth via the star formation rate (SFR) which can in turn be compared to the observed star formation rate providing an additional and independent validation test to the model. Using an empirical quenching model derived from observations the full distributions of specific star formation rate (sSFR) can be modelled and compared to observations. This highly constrained and deliberate modelling is ideal for understanding the effects of individual theories in galactic astrophysics which would be very difficult to constrain via alternative, more traditional, approaches.

2.2.2 Using STEEL to check the self consistency of fitting models and cosmological models.

STEEL can be used to check self consistency of fitting models and observations. The follow steps provide an illustrative example for a consistency check using examples that I have had success with in my PhD:

- Observations are taken by a given mission e.g. Euclid, and a cosmology e.g. Planck is chosen to analyse the results in.
- Fitting models are used to derive galaxy properties over multiple epochs e.g. Stellar masses, Star formation rate, pair fractions...
- STEEL is run using a statistical dark matter accretion history calibrated on the chosen cosmology and the derived galaxy properties which are used to generate a SMHM relation and thus to assign galaxies of a given stellar mass/SFR to host dark matter haloes.
- STEEL then tests whether the evolution implied by the derived galaxy properties are consistent with the evolution expected from ΛCDM hierarchical assembly.
- There are multiple 'tiers' of results from STEEL: For example at its most basic the rate of satellite accretion can be compared to the rate of galaxy growth if this is not consistent (i.e there is more accretion than growth) it is reasonable to reject either the fits or ΛCDM cosmology. More complex modelling can predict star formation rates that would be required to maintain the galaxy growth and these compared to the observed star formation rate, should these disagree it is likely that our understanding of star formation processes or the fitting of star formation is incorrect.

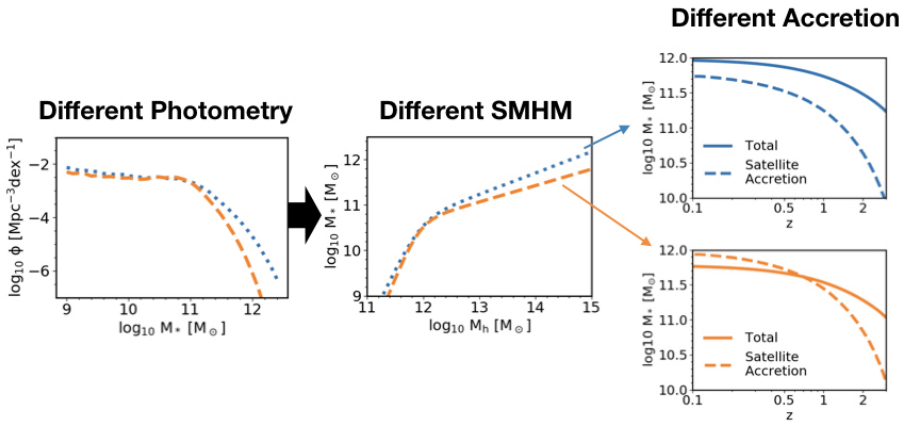


Figure 2: Schematic cartoon of how observational differences propagate into modelled inconsistencies.

Figure 2 is a simplification of the derivation of the satellite accretion vs. growth rate described in Section 2.2.1. In this example I show how a change in the observations caused by a photometric (fitting) choice propagates into an nonphysical satellite accretion that exceeds the central galaxy growth rate. This is an example of the lowest 'tier' analysis that can be used to check for consistency in data fitting models and ΛCDM assembly models, but it already rejects with good certainty traditional stellar mass fitting models (Grylls2019b, submitted). Advancing this technique to include more 'tiers' and tailoring it

to additional galaxy properties of interest STEEL will bring theoretical understanding closer to the data we use to validate theory. This will simultaneously: increase the reliability of observations, constrain cosmological models, and advance understanding of galaxy assembly theory.

2.2.3 Galaxy modelling with STEEL

One of the major questions in the field of galactic astrophysics is how galaxies become quenched, i.e. what causes a galaxy to cease converting gas into stars. Several possible solutions have been discussed including AGN feedback, galaxy mergers, and/or, quenching due to inefficiency of gas supply though dark matter haloes above a given mass. Semi-Analytic and Hydrodynamical simulations obscure physics such as this due to the degeneracy of tuning parameters.

For example, if one were to attempt to test the connection between AGN feedback and star formation rate in a stable model without AGN it is likely that introduction of AGN will result in far more than just the re-tuning of starformation associated parameters. Due to the change in tuning it then renders the AGN and non-AGN models hard to compare fairly obscuring the connection between AGN and starformation. A Semi-Empirical model is better suited to solving these problems as they are capable of running with only the essential physics simply re-initialising essential galaxy properties to be consistent with observations at each time-step. In this way only the physical processes under investigation contribute to the results of the simulation making it extremely flexible, fast, and transparent.

An illustrative example of this technique is a simple model of galaxy growth in a central dark matter halo. At each time-step the central halo is assigned stellar mass using the stellar-mass-halo-mass relation. Introducing galaxy mergers by following the merging dark matter substructures the mass available to the central galaxy from accretion is calculated. If the accreted mass is sufficient to grow the galaxy to the size predicted by the stellar-mass-halo-mass-relation at later times then it is reasonable to argue that this galaxy grows mainly though mass accretion. If the accretion is too high then there must be more physics involved in the satellites i.e. stripping, or the merging model is wrong. If the accretion is too low to explain the growth of the galaxy then there is some additional mass growth mechanism i.e. star formation, or the merging model is wrong. With an empirical model understanding the physics and testing theories is much clearer.

Considering the above semi-empirical modelling example it becomes evident that higher-quality data will allow for more constraining power. For example, good star formation rate data could be used as a further input in the model to aid the growth of the galaxy, thus removing a degree of freedom. For an empirical model more data simply adds power, for a semi-analytic or hydrodynamical model it adds a constraint that if in disagreement with the model requires significant remodelling.

Open galaxy modelling questions that I will constrain during this fellowship are:

- Compare and contrast the effectiveness of proposed mechanisms for satellite galaxy quenching.
  - Central AGN feedback causing ram pressure stripping on satellites.
  - Stripping of the satellite gas reservoir by the halo environment.
- Understand the relative contributions of different processes to central galaxy quenching.
  - Mergers causing morphological change associated with the end of starformation.
  - AGN feedback suppressing star formation directly or coupling with the halo environment.
- Distinguish which populations of galaxies are transformed into lenticulars and by which mechanism(s).

With each of these investigations the modelling process is to take theoretical models, often interpreted from observations or small high resolution hydrodynamic simulations, and apply it to a statistical model to understand how these effects would manifest in the full galaxy population.

2.3 Impact of Research

In addition to being a novel approach, STEEL has also been devolved with several upcoming large surveys in mind. EUCLID, LSST, eROSITA, Athena, and JWST each have the capability to greatly impact our theoretical understanding of galaxy formation and cosmology. However, as shown with my previous results without accurate theoretical modelling to inform the data fitting models, these surveys risk the same inconsistencies as previous missions, sacrificing potential impact at great fiscal cost to the scientific community. In addition, fitting/cosmologies have previously been corrected with overly simplistic factors. For example to convert from a Chabrier (Chabrier, 2003) initial mass function to a Salpeter (Salpeter, 1955) inital mass function multiplicative factor of 0.63 is used and cosmological corrections use only a factor called 'little h' (Croton, 2013). I have shown how the propagation of inputs create a requirement for more subtle corrections in shape and normalisation. STEEL can be used to create cosmology/fit corrections that capture these subtleties and allow years of old data to be used with the high comparative quality that STEEL will provide to new surveys. Current models struggle to produce the volumes currently observed so are unlikely to keep up with the larger volumes soon to be observed where as for a statistical model this is a non-issue. STEEL is ideally and uniquely suited for the future of galactic modelling and the advancement of understanding the formation and nature of our Universe.

2.4 Research Plan

For this project to have maximum impact it will require strong early collaboration with groups such as the Euclid consortium. STEEL is already known by several influential empirical modellers. My research group in Southampton is heavily involved in the Euclid collaboration for which STEEL will represent an invaluable tool to produce reliable "mock" test galaxy catalogues and to properly interpret the data. This is an imminent vital task given that Euclid is currently scheduled for 2022.

The first goal is to use STEEL to make a tool that checks fitting model and cosmological model consistency. Working with contacts from the Euclid collaboration this can be prepared with the flagship mocks such that the tool is ready for launch. Following launch we will have access to consistency checked surveys which will greatly increase the empirical power of the model. With this highly constrained tests of galaxy formation theories can be carried out with STEEL to place constraints on the quenching of galaxies and other open questions.

To ensure maximum impact via collaboration and awareness during the first year I will organise a workshop including modellers and observers. The aims of this workshop will be:

- Publicise the 'fitting model'-'cosmological model' inconsistencies. Make observers aware of how this should influence the way they approach fitting. Make modellers aware of how the data should be considered when comparing to models.
- Gain support from those people working on large surveys to adopt STEEL as a method to help develop consistent fits.
- Build discourse between observers and theorists about the open problems and the systematics present in both observation and modelling that are barriers to making clear theoretical progress.

3 Professional Development

The University of Southampton provides many courses to ensure the continued professional development of their staff. Attending the following courses during the first half of the fellowship will enhance my ability to produce and communicate my science.

- Optimising your time: With an ambitious project such as this it will be important to me to improve my already strong self-management skills to ensure timely progress.

- Successful communication: I will continue to improve my communication skills to ensure an efficient transfer of knowledge of the core modelling concepts behind STEEL create maximum impact.
- Peer Review: The process of peer review is essential to develop outstanding skills in reading and writing high-quality papers.
- Quality Papers: Papers are the foundation of a good career and are the core material by which work is judged and recognised. With continual development I will further improve my papers to ensure my techniques are visible and accessible to the community.
- Writing Science for the public: The scientific community are funded by the public and have a duty to report their findings in a timely and engaging way. This course will allow me to ensure my research is well received. To provide continual relatable content I will make a monthly blog on what I am working on pitched toward general public.

In addition to these courses I also have strong links to the Research Software Engineering community keeping me up to date with the latest in best computational practice. My models and codes will be efficient as possible but also well documented, this means my work is accessible to the modelling community.

4 Public Engagement

4.1 Track Record

I have a wealth of experience in public engagement. Since starting University I have provided over 100 hours of engagement activities which include:

- Bath TAP: General science focusing on forces. Age range 5-12 years.
- BRLSI: The 'Bright Spark' workshops focusing on electromagnetism. Age range 7-14 years.
- Planetrella: Talk and a live demo of how plasma is contained by magnetic fields used to explain the relevance of the Aurora to our daily lives and the history of their discovery. Age range 4-90 years.

In addition to this I have a teaching qualification in Mathematics for the 11-18 age range so I am comfortable disseminating ideas to young audiences. In addition to my outreach work I have also spent over 50 hours volunteering in a school for autistic children so have familiarity working with less able people.

4.2 Plans and experience engaging audiences

4.3 Activities

STEEL is a lightweight model that can be configured to run very quickly. We can use this to create demos where people interact with the actual scientific code. The following are examples of demos that I expect to develop to be able to to engage the general public with the beauty of galaxy modelling.

- Use STEEL combined with the outreach tool ASTERA, an interactive cosmological 3D galaxy visualisation program, to create image mocks of galaxies in the Universe at any epoch given different stellar-mass-halo-mass relations. Asking the public to change the stellar-mass-halo-mass parameter they can try and match the look of real clusters or create 'extreme universes'.
- One of the joys of empirical modelling is that some of the key results can be approximated very simply. With the rise in schools teaching basic programming this should make GCSE and A-Level students be able to code and access the results with minimal extra intervention. By providing simplified halo and galaxy catalogues it will be possible to create a workshop to teach the basic principles of empirical modelling.

4.4 Training

In addition to the professional development the University of Southampton offer outreach specific courses. Before beginning the public engagement activities I will take these three courses to ensure I am up to date with effective engagement strategies.

- Introduction to Public Engagement with Research
- Engagement with schools
- Introduction to one to many engagement.

References

Asquith R., et al., 2018, http://dx.doi.org/10.1093/mnras/sty1870 Monthly Notices of the Royal Astronomical Society, 480, 1197

Bull P., et al., 2016, Beyond ΛCDM: Problems, solutions, and the road ahead, http://dx.doi.org/10.1016/j.dark.2016.02.001 doi:10.1016/j.dark.2016.02.001, http://arxiv.org/abs/1502.01589 http://dx.doi.org/10.1051/0004-6361/201525830 http://www.aanda.org/10.1051/0004-6361/201525830

Chabrier G., 2003, http://dx.doi.org/10.1086/376392 Publications of the Astronomical Society of the Pacific, 115, 763

Croton D. J., 2013, Publications of the Astronomical Society of Australia

Grylls P. J., Shankar F., Zanisi L., Bernardi M., 2019, http://dx.doi.org/10.1093/mnras/sty3281 MNRAS, 483, 2506

Lapi A., Cavaliere A., 2011, http://dx.doi.org/10.1155/2011/903429 Advances in Astronomy, 2011, 1

Parkinson H., Cole S., Helly J., 2008, http://dx.doi.org/10.1111/j.1365-2966.2007.12517.x Monthly Notices of the Royal Astronomical Society, 383, 557

Planck Collaboration et al., 2015, http://dx.doi.org/10.1051/0004-6361/201525830 Astronomy & Astrophysics, 594

Salpeter E. E., 1955, http://dx.doi.org/10.1086/145971 The Astrophysical Journal, 121, 161

Wang T., et al., 2016, http://dx.doi.org/10.3847/0004-637X/828/1/56 The Astrophysical Journal, 828

Hawking Fellowship: Pathways to Impact

1 Research Impacts

My research project is expected to have two major impacts in the scientific/research community. Firstly, using STEEL I intend to add constraints to long standing questions in galaxy formation such as:

- The connection between galaxy mergers, AGN, and, star formation in central galaxies.
- Constrain the contributions of effects such as ram pressure stripping, strangulation, and harassment on the quenching of satellite galaxies in the group/cluster environment.

Secondly, using STEEL as a tool alongside galaxy fitting routines I will add reliability and consistency to the outcomes of upcoming major surveys such as Euclid. The methods though which I will ensure the academic impact of this project are as follows:

- Use contacts within the Euclid collaboration to ensure STEEL is used as part of the mock catalogue creation and used within the data pipeline concerning fitting models.
- Organise a workshop to publicise the motivation behind the fellowship. Inviting key figures in the modelling community as well as key researchers from upcoming major surveys will ensure the project has support and visibility.
- The results of this fellowship will create numerous peer-reviewed manuscripts and presentations at conferences both nationally and internationally.

2 Outreach Impacts

The aims of outreach with STEEL will be twofold:

- Promote interest in galactic cosmology in the general public and what it can teach us about the universe we live in.
- Give A-Level students hands-on experience with empirical modelling and real galaxy data to make them enthusiastic about pursuing Physics at University and a career in research.

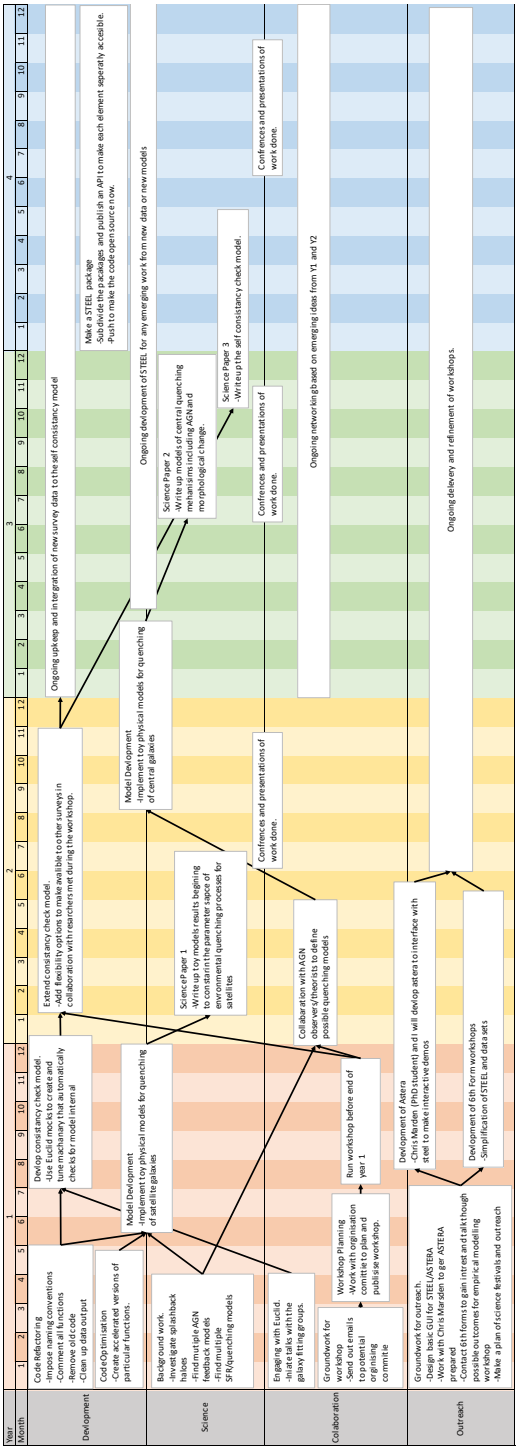
I have two planned public engagement projects for use with STEEL:

- Create an engaging demonstration by combing Astera, a cosmological galaxy visualisation code developed in the group at Southampton, with the outputs of STEEL. By use of careful prepossessing and accelerations of these codes it will be possible to make an 'interactive' Universe where it is possible to directly see the effects of changing cosmological parameters or assumptions.
- Create a empirical modelling workshop using a reduced data set and a simplified version of STEEL. The workshop will be aimed to engage the 16-18 age range, and show young students what actual research looks like.

3 Economic Impacts

The cost of any observational mission is huge. For example, Euclid has a budget of €500 million, and JWST is reported to cost of order $10 billion. The types of flaws in observational fitting found in my work substantially reduce the abilities of missions such as these to achieve their science goals of precision cosmology and limit our theoretical understanding of the universe. The work I will undertake during this fellowship will ensure these surveys are self consistent in observational fitting models and cosmological models. In addition to this STEEL will provide corrections to previous missions to unbias the data they provided allowing it to also be used for precision cosmology. Left unaddressed, the issues I present will reduce the impact of the UK's (and worldwide scientific community's) current scientific investment and could potentially lead to misplaced future investments.

We are entering the era of 'big data'. however without suitable analysis techniques we seriously risk not to take proper advantage of incoming massive data sets. STEEL is the ideal model to coexist with astronomical 'big data' and will ensure the value of the UK's investment in theoretical astronomy.


Year
Month
Development
Code Refactoring
-Improve coding conventions
-Comment all functions
-Remove old code
-Clean up data output
Code Optimisation
-Create accelerated versions of particular functions.
Develop consistency check model.
-Use code of this to create and tune machanism that automatically checks for model internal
Extend consistancy check model.
-Add flexibility options to make available to other surveys in collaboration with researchers met during the workshop.
Ongoing upkeep and intergration of new survey data to the self consistancy model
Make a STEEL package
-Sub divide the packages and publish an API to make each element seperatly accesible.
-Push to make the code open source enow.
Science
Background work.
-Investigate splashback haloes
-Find multiple AGN feeback models
-Find multiple SFR/quenching models
Model Development
-Implement toy physical models for quenching of satellite galaxies
Model Development
-Implement toy physical models for quenching of central galaxies
Science Paper 1
-Write up toy models results begining to constarin the parameter sapce of environmental quenching processes for satellites
Science Paper 2
-Write up models of central quenching mechanism and effect on SMHM and morphological change
Science Paper 3
-Write up the self consistancy check model.
Ongoing development of STEEL for any emerging work from new data or new models
Collaboration
Engaging with Euclid.
-Initial talks with the galaxy clustering groups.
Groundwork for workshop
-Send out emails to potential organising committe
Workshop Planning
-Work with orgnisation committe to plan and publisise workshop.
Run workshop before end of year 1
Collaboration with AGN observer/theorists to define possible quenching models
Confrences and presentations of work done.
Confrences and presentations of work done.
Confrences and presentations of work done.
Ongoing networking based on emerging ideas from Y1 and Y2
Outreach
-Design basic GUI for STEEL/ASTERA
-Work with Chris Madden to get ASTERA prepared
-Contact 6th forms to gain intrest and talk though possible outcomes for empirical modelling workshop
-Make a plan of science festivals and outreach
Development of Astera
-Chris Marken (PhD student) and I will develop astera to interface with steel to make interactive demos
Development of 6th Form workshops
-Simplification of STEEL and data sets
Ongoing delevery and refinement of workshops.

Bibliography

[1] Rachel S Somerville, GergÃűi Popping, and Scott C Trager. Star formation in semi-analytic galaxy formation models with multiphase gas. *Monthly Notices of the Royal Astronomical Society*, 453(4):4337–4367, 2015. ISSN 13652966. doi: 10.1093/mnras/stv1877.

[2] Aldo Rodríguez-Puebla, Joel R. Primack, Vladimir Avila-Reese, and S. M. Faber. Constraining the galaxy-halo connection over the last 13.3 Gyr: Star formation histories, galaxy mergers and structural properties. *Monthly Notices of the Royal Astronomical Society*, 470(1):651–687, 2017. ISSN 13652966. doi: 10.1093/mnras/stx1172. URL https://academic.oup.com/mnras/article-abstract/470/1/651/3820932.

[3] James Webb Space Telescope (WEBB/JWST) NASA. URL https://www.jwst.nasa.gov/%0Ahttps://jwst.nasa.gov/index.html.

[4] Carl J Mundy, Christopher J Conselice, Kenneth J. Duncan, Omar Almaini, Boris Häußler, and William G Hartley. A consistent measure of the merger histories of massive galaxies using close-pair statistics I: Major mergers at $z < 3.5$. *MNRAS*, 470(3):3507–3531, 5 2017. doi: 10.1093/mnras/stx1238. URL http://www.gama-survey.org/dr2/schema/table.php?id=168https://arxiv.org/pdf/1705.07986.pdfhttp://arxiv.org/abs/1705.07986http://dx.doi.org/10.1093/mnras/stx1238.

[5] M. Cook, A. Lapi, and G. L. Granato. Two-phase galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 397(1):534–547, 7 2009. ISSN 00358711. doi: 10.1111/j.1365-2966.2009.14962.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/397/1/10.1111/j.1365-2966.2009.14962.x/3/mnras0397-0534.pdf?Expires=1495894118&Signature=Qu0LSHFwev4YOCOouCsAe2Y2bCuP~C3KW~~GP0cO~-FJWYIv-qd1SbWQOpRkj8fqtBtJnrFDhkuSPtgiH

[6] Peter Behroozi, Risa H Wechsler, Andrew P Hearin, and Charlie Conroy. UniverseMachine: The correlation between galaxy growth and dark matter halo

assembly from z = 0âĹŠ10. *Monthly Notices of the Royal Astronomical Society*, 488(3):3143–3194, 9 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1182. URL http://arxiv.org/abs/1806.07893http://dx.doi.org/10.1093/mnras/stz1182https://academic.oup.com/mnras/article/488/3/3143/5484868.

[7] Dylan Nelson, Annalisa Pillepich, Volker Springel, RÃijdiger Pakmor, Rainer Weinberger, Shy Genel, Paul Torrey, Mark Vogelsberger, Federico Marinacci, and Lars Hernquist. First results from the TNG50 simulation: Galactic outflows driven by supernovae and black hole feedback. *Monthly Notices of the Royal Astronomical Society*, 490(3):3234–3261, 2 2019. ISSN 13652966. doi: 10.1093/mnras/stz2306. URL https://academic.oup.com/mnras/advance-article/doi/10.1093/mnras/stz2306/5556547http://arxiv.org/abs/1902.05554http://dx.doi.org/10.1093/mnras/stz2306.

[8] Benjamin P Moster, Thorsten Naab, and Simon D M White. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 428(4):3121–3138, 2 2013. ISSN 0035-8711. doi: 10.1093/mnras/sts261. URL http://academic.oup.com/mnras/article/428/4/3121/997374/Galactic-star-formation-and-accretion-histories.

[9] Andrew R Wetzel, Jeremy L Tinker, Charlie Conroy, and Frank C Van Den Bosch. Galaxy evolution in groups and clusters: satellite star formation histories and quenching time-scales in a hierarchical Universe. *MNRAS*, 432 (1):336–358, 2013. doi: 10.1093/mnras/stt469. URL https://watermark.silverchair.com/stt469.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAb0wggG5BgkqhkiG9w0BBwagggGqMIIBpgIBADCCAZ8GCSqGSIb3DQEHATAeBglghkgBZQMGctEuQtRDAgEQgIIBcHe8UzgtFyDjRe7SMuPQqJGtQQeWyXOXSQ3sWVLpEhiS9VMO.

[10] Sean P Fillingham, Michael C Cooper, Andrew B Pace, Michael Boylan-Kolchin, James S Bullock, Shea Garrison-Kimmel, and Coral Wheeler. Under pressure: Quenching star formation in low-mass satellite galaxies via stripping. *Monthly Notices of the Royal Astronomical Society*, 463(2):1916–1928, 2016. ISSN 13652966. doi: 10.1093/mnras/stw2131. URL https://arxiv.org/pdf/1606.07810.pdf.

[11] Adam R Tomczak, Ryan F Quadri, Kim-Vy H Tran, Ivo Labbé, Caroline M S Straatman, Casey Papovich, Karl Glazebrook, Rebecca Allen, Gabreil B Brammer, Michael Cowley, Mark Dickinson, David Elbaz, Hanae Inami, Glenn G Kacprzak, Glenn E Morrison, Themiya Nanayakkara, S Eric Persson, Glen A Rees, Brett Salmon, Corentin Schreiber, Lee R Spitler, and Katherine E Whitaker. THE SFRâĂŞM* RELATION AND EMPIRICAL STAR FORMATION HISTORIES

FROM ZFOURGE AT 0.5 < z < 4. *The Astrophysical Journal*, 817:118, 2016. doi: 10.3847/0004-637X/817/2/118.

[12] Tao Wang, David Elbaz, Emanuele Daddi, Alexis Finoguenov, Daizhong Liu, Corentin Schreiber, Sergio Martín, Veronica Strazzullo, Francesco Valentino, Remco van der Burg, Anita Zanella, Laure Ciesla, Raphael Gobat, Amandine Le Brun, Maurilio Pannella, Mark Sargent, Xinwen Shu, Qinghua Tan, Nico Cappelluti, and Yanxia Li. DISCOVERY OF A GALAXY CLUSTER WITH A VIOLENTLY STARBURSTING CORE AT z = 2.506. *The Astrophysical Journal*, 828(1):56, 8 2016. ISSN 1538-4357. doi: 10.3847/0004-637X/828/1/56. URL http://stacks.iop.org/0004-637X/828/i=1/a=56?key=crossref.91f76f9ada918a13ca710b56f6e1fe14.

[13] Z L Wen and J L Han. A sample of 1959 massive galaxy clusters at high redshifts. *Monthly Notices of the Royal Astronomical Society*, 481(3):4158–4168, 9 2018. ISSN 13652966. doi: 10.1093/mnras/sty2533.

[14] Joel Leja, Benjamin D Johnson, Charlie Conroy, Pieter van Dokkum, Joshua S Speagle, Gabriel Brammer, Ivelina Momcheva, Rosalind Skelton, Katherine E Whitaker, Marijn Franx, and Erica J Nelson. An Older, More Quiescent Universe from Panchromatic SED Fitting of the 3D- HST Survey. *The Astrophysical Journal*, 877(2):140, 6 2019. ISSN 1538-4357. doi: 10.3847/1538-4357/ab1d5a. URL https://arxiv.org/pdf/1812.05608.pdfhttp://arxiv.org/abs/1812.05608http://dx.doi.org/10.3847/1538-4357/ab1d5ahttps://iopscience.iop.org/article/10.3847/1538-4357/ab1d5a.

[15] Piero Madau and Mark Dickinson. Cosmic Star-Formation History. *Annual Review of Astronomy and Astrophysics*, 52(1):415–486, 2014. ISSN 0066-4146. doi: 10.1146/annurev-astro-081811-125615. URL https://arxiv.org/pdf/1001.0061.pdfhttp://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010arXiv1001.0061R&link_type=ABSTRACT%5Cnpapers://ee00755c-a478-4d4e-a50b-ef01ee7b9957/Paper/p4275http://arxiv.org/abs/1403.0007v3%5Cnhttp://www.annualreviews.o.

[16] Joel Leja, Joshua S Speagle, Benjamin D Johnson, Charlie Conroy, Pieter Van Dokkum, and Marijn Franx. A New Census of the 0.2 < z < 3.0 Universe, Part I: The Stellar Mass Function. Technical report, 2020.

[17] Benjamin P. Moster, Thorsten Naab, and Simon D. M. White. emerge âĂŞ an empirical model for the formation of galaxies since z âĹij 10. *Monthly Notices of the Royal Astronomical Society*, 477(2):1822–1852, 6 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty655. URL http://arxiv.org/abs/1705.05373http:

//dx.doi.org/10.1093/mnras/sty655https://academic.oup.com/mnras/article/477/2/1822/4939290.

[18] N Menci, M Gatti, F Fiore, and A Lamastra. Triggering active galactic nuclei in hierarchical galaxy formation: disk instability vs. interactions. *Astronomy & Astrophysics*, 569(A37):A37, 9 2014. ISSN 0004-6361. doi: 10.1051/0004-6361/201424217. URL http://www.aanda.orghttp://www.aanda.org/10.1051/0004-6361/201424217.

[19] Philip J Grylls, F Shankar, L Zanisi, and M Bernardi. A statistical semi-empirical model: satellite galaxies in groups and clusters. *Monthly Notices of the Royal Astronomical Society*, 483(2):2506–2523, 2 2019. ISSN 0035-8711. doi: 10.1093/mnras/sty3281. URL https://academic.oup.com/mnras/article-abstract/483/2/2506/5228747https://academic.oup.com/mnras/article/483/2/2506/5228747.

[20] Philip J. Grylls, F. Shankar, J. Leja, N. Menci, B. Moster, P. Behroozi, and L. Zanisi. Predicting fully self-consistent satellite richness, galaxy growth, and star formation rates from the STatistical sEmi-Empirical modeL STEEL. *Monthly Notices of the Royal Astronomical Society*, 491(1):634–654, 2020. ISSN 13652966. doi: 10.1093/mnras/stz2956. URL https://academic.oup.com/mnras/article-abstract/491/1/634/5602609.

[21] Philip J. Grylls, F. Shankar, and C. Conselice. The significant effects of stellar mass estimation on galaxy pair fractions. *MNRAS*, 000:1–13, 1 2020. URL http://arxiv.org/abs/2001.06017.

[22] Rachel S. Somerville, Philip F. Hopkins, Thomas J. Cox, Brant E. Robertson, and Lars Hernquist. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 391 (2):481–506, 2008. ISSN 00358711. doi: 10.1111/j.1365-2966.2008.13805.x.

[23] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1): 2, 12 2019. ISSN 2197-7909. doi: 10.1186/s40668-019-0028-x. URL www.tng-project.org/datahttps://arxiv.org/abs/1812.05609http://arxiv.org/abs/1812.05609https://comp-astrophys-cosmol.springeropen.com/articles/10.1186/s40668-019-0028-x.

[24] Edwin Hubble. Extra-galactic nebulae. *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 324:1–49, 1926.

[25] Edwin Hubble. The classification of spiral nebulae. *The Observatory*, 50:276–281, 1927.

[26] Erik Holmberg. On the Clustering Tendencies among the Nebulae. II. a Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure. *The Astrophysical Journal*, 94(633):385, 11 1941. ISSN 0004-637X. doi: 10.1086/144344. URL http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1941ApJ....94..385H&data_type=PDF_HIGH&whole_paper=YES&type=PRINTER&filetype=.pdfhttp://adsabs.harvard.edu/doi/10.1086/144344.

[27] Stuart McAlpine, John C. Helly, Matthieu Schaller, James W. Trayford, Yan Qu, Michelle Furlong, Richard G. Bower, Robert A. Crain, Joop Schaye, Tom Theuns, Claudio Dalla Vecchia, Carlos S. Frenk, Ian G. McCarthy, Adrian Jenkins, Yetli Rosas-Guevara, Simon D. M. White, Maarten Baes, Peter Camps, and Gerard Lemson. The EAGLE simulations of galaxy formation: public release of halo and galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 450(2): 1937–1961, 10 2015. ISSN 13652966. doi: 10.1016/j.ascom.2016.02.004.

[28] Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, RÃijdiger Pakmor, Paul Torrey, Rainer Weinberger, Shy Genel, Jill P Naiman, Federico Marinacci, and Mark Vogelsberger. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, 475(1):648–675, 3 2018. ISSN 0035-8711. doi: 10.1093/mnras/stx3112. URL https://arxiv.org/pdf/1707.03406.pdfhttps://academic.oup.com/mnras/article/475/1/648/4683271.

[29] H J Mo, Shude Mao, and Simon D M White. The formation of galactic discs. *Monthly Notices of the Royal Astronomical Society*, 295(2):319–336, 4 1998. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1998.01227.x. URL http://arxiv.org/abs/astro-ph/9707093http://dx.doi.org/10.1046/j.1365-8711.1998.01227.xhttps://academic.oup.com/mnras/article-lookup/doi/10.1046/j.1365-8711.1998.01227.x.

[30] Gabriella De Lucia, Volker Springel, S. D. M. White, Darren Croton, and Guinevere Kauffmann. The formation history of elliptical galaxies. *Monthly Notices of the Royal Astronomical Society*, 366(2):499–509, 2 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2005.09879.x. URL https://academic.oup.com/mnras/

article-abstract/366/2/499/1215246https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2005.09879.x.

[31] Qi Guo, Simon White, Michael Boylan-Kolchin, Gabriella De Lucia, Guinevere Kauffmann, Gerard Lemson, Cheng Li, Volker Springel, and Simone Weinmann. From dwarf spheroidals to cD galaxies: Simulating the galaxy population in a ΛCDM cosmology. *Monthly Notices of the Royal Astronomical Society*, 413(1): 101–131, 2011. ISSN 00358711. doi: 10.1111/j.1365-2966.2010.18114.x.

[32] Philip F. Hopkins, Kevin Bundy, Darren Croton, Lars Hernquist, Dusan Keres, Sadegh Khochfar, Kyle Stewart, Andrew Wetzel, and Joshua D. Younger. MERGERS AND BULGE FORMATION IN ΛCDM: WHICH MERGERS MATTER? *The Astrophysical Journal*, 715(1):202–229, 5 2010. ISSN 0004-637X. doi: 10.1088/0004-637X/715/1/202. URL http://iopscience.iop.org/article/10.1088/0004-637X/715/1/202/pdfhttp://stacks.iop.org/0004-637X/715/i=1/a=202?key=crossref.d956c23377616fd48555c7b449c86600http://www.cfa.harvard.edu/~phopkins/Site/mergercalc.

[33] Jesus Zavala, Vladimir Avila-Reese, Claudio Firmani, and Michael Boylan-Kolchin. The growth of galactic bulges through mergers in Λ CDM haloes revisited - I. Present-day properties. *Monthly Notices of the Royal Astronomical Society*, 427(2):1503–1516, 2012. ISSN 00358711. doi: 10.1111/j.1365-2966.2012.22100.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/427/2/10.1111/j.1365-2966.2012.22100.x/2/427-2-1503.pdf?Expires=1490547069&Signature=Y2lMxvaxNY9p-Yp9OHSq0hzRNMjMeSuNHkg5NtX0GLNLNHBh3FYuQqYcXEyz3XpsVxkmXOIRcys37hUsOgUVtNuh

[34] Francesco Shankar, Simona Mei, Marc Huertas-Company, Jorge Moreno, Fabio Fontanot, Pierluigi Monaco, Mariangela Bernardi, Andrea Cattaneo, Ravi Sheth, Rossella Licitra, Lauriane Delaye, and Anand Raichoor. Environmental dependence of bulge-dominated galaxy sizes in hierarchical models of galaxy formation. Comparison with the local Universe. *Monthly Notices of the Royal Astronomical Society*, 439(4):3189–3212, 3 2014. ISSN 0035-8711. doi: 10.1093/mnras/stt2470. URL http://mnras.oxfordjournals.org/cgi/doi/10.1093/mnras/stt2470.

[35] F Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.

[36] M. S. Roberts and A. H. Rots. Comparison of Rotation Curves of Different Galaxy Types. *Astronomy and Astrophysics*, 26:483–485, 1973. ISSN 0004-6361. URL http://adsabs.harvard.edu/abs/1973A%26A....26..483R.

[37] Brad K. Gibson and Jeremy R. Mould. The Chemical Residue of a White DwarfâĂŞdominated Galactic Halo. *The Astrophysical Journal*, 482(1):98–103, 6 1997. ISSN 0004-637X. doi: 10.1086/304140. URL https://iopscience.iop.org/article/10.1086/304140.

[38] Von A Einstein. Sitzung der physikalisch-mathematischen Klasse vom 8. Februar 1917 Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. Technical report, 1917.

[39] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R.-R. Chary, H. C. Chiang, J. Chluba, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F.-X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, M. Farhang, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, T. Ghosh, M. Giard, Y. Giraud-Héraud, E. Giusarma, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marchini, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J.-B. Melin, L. Mendes, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge,

F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J.-L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, B. Rouillé dâĂŹOrfeuil, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, N. Said, V. Salvatelli, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. *Astronomy & Astrophysics*, 594(A13):A13, 10 2016. ISSN 0004-6361. doi: 10.1051/0004-6361/201525830. URL http://arxiv.org/abs/1502.01589http://dx.doi.org/10.1051/0004-6361/201525830http://www.aanda.org/10.1051/0004-6361/201525830http://www.esa.int/Planck.

[40] Michael Boylan-Kolchin, Volker Springel, Simon D.M. White, Adrian Jenkins, and Gerard Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. *Monthly Notices of the Royal Astronomical Society*, 398(3):1150–1164, 2009. ISSN 00358711. doi: 10.1111/j.1365-2966.2009.15191.x. URL https://academic.oup.com/mnras/article-abstract/398/3/1150/1261296.

[41] Anatoly Klypin, Gustavo Yepes, Stefan Gottlöber, Francisco Prada, and Steffen Heß. Multidark simulations: The story of dark matter halo concentrations and density profiles. *Monthly Notices of the Royal Astronomical Society*, 457(4):4340–4359, 2016. ISSN 13652966. doi: 10.1093/mnras/stw248.

[42] Anatoly Klypin and Jon Holtzman. Particle-Mesh code for cosmological simulations. *arXiv e-prints*, 1997. URL http://astro.nmsu.edu/http://arxiv.org/abs/astro-ph/9712217.

[43] Marc Davis, George Efstathiou, Carlos S Frenk, and Simon D M White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371, 5 1985. ISSN 0004-637X. doi: 10.1086/163168. URL http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1985ApJ...292..371D&data_type=PDF_HIGH&whole_paper=YES&type=PRINTER&filetype=.pdfhttp://adsabs.harvard.edu/doi/10.1086/163168.

[44] Peter S Behroozi, Risa H Wechsler, and Hao-Yi Yi Wu. The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophysical Journal*, 762(2):109, 10 2011. ISSN 15384357. doi: 10.1088/0004-637X/762/2/109. URL http://code.google.com/p/consistent-trees.Ourtreesandcatalogsarepubliclyavailableathttp://hipacc.ucsc.edu/Bolshoi/.http://code.google.com/p/rockstar.http://stacks.iop.org/0004-637X/762/i=2/a=109?key=crossref.6bf709f598ac69743b201969e4dff7cehttp://arxiv.

[45] Stuart P.D. Gill, Alexander Kneb, and Brad K. Gibson. The evolution of substructure -I. A new identification method. *Monthly Notices of the Royal Astronomical Society*, 351(2):399–409, 6 2004. ISSN 00358711. doi: 10.1111/j.1365-2966.2004.07786.x.

[46] Frank C. van den Bosch and Go Ogiya. Dark matter substructure in numerical simulations: A tale of discreteness noise, runaway instabilities, and artificial disruption. *Monthly Notices of the Royal Astronomical Society*, 475(3):4066–4087, 1 2018. ISSN 13652966. doi: 10.1093/MNRAS/STY084. URL https://arxiv.org/pdf/1801.05427.pdfhttp://arxiv.org/abs/1801.05427http://dx.doi.org/10.1093/mnras/sty084.

[47] Matthias Bartelmann and Peter Schneider. Weak gravitational lensing, 1 2001. ISSN 03701573.

[48] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *The Astrophysical Journal*, 187:425, 2 1974. ISSN 0004-637X. doi: 10.1086/152650. URL http://adsabs.harvard.edu/doi/10.1086/152650.

[49] Rodrigo Voivodic, Marcos Lima, and Luis Raul Abramo. Excursion Set Halos – ExSHalos: A New Parameter Free Method for Fast Generation of Halo Catalogues. 6 2019. URL http://arxiv.org/abs/1906.06630.

[50] Shaun Cole, Cedric Lacey, Carlton Baugh, and Carlos Frenk. Hierarchical Galaxy Formation. *MNRAS*, 319(1):168–204, 2000. ISSN 00358711. doi: 10.1046/j.1365-8711.2000.03879.x. URL http://arxiv.org/abs/astro-ph/0007281%0Ahttp://dx.doi.org/10.1046/j.1365-8711.2000.03879.x.

[51] Hannah Parkinson, Shaun Cole, and John Helly. Generating dark matter halo merger trees. *Monthly Notices of the Royal Astronomical Society*, 383(2):557–564, 2008. ISSN 00358711. doi: 10.1111/j.1365-2966.2007.12517.x.

[52] Volker Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Monthly Notices of the Royal Astronomical Society*, 401(2):791–851, 2010. ISSN 00358711. doi: 10.1111/j.1365-2966.2009.15715.x. URL https://arxiv.org/pdf/0901.4107.pdf.

[53] Volker Springel and Lars Hernquist. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. *Monthly Notices of the Royal Astronomical Society*, 339(2):289–311, 2 2003. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2003.06206.x. URL https://arxiv.org/pdf/astro-ph/0206393.pdfhttp://arxiv.org/abs/astro-ph/0206393%0Ahttp://dx.doi.org/10.1046/j.1365-8711.2003.06206.xhttps://academic.oup.com/mnras/article-lookup/doi/10.1046/j.1365-8711.2003.06206.x.

[54] Philip F Hopkins, Andrew Wetzel, DuÅąan Kereš, Claude AndrÃĺ Faucher-Giguère, Eliot Quataert, Michael Boylan-Kolchin, Norman Murray, Christopher C Hayward, Shea Garrison-Kimmel, Cameron Hummels, Robert Feldmann, Paul Torrey, Xiangcheng Ma, Daniel Anglés-Alcázar, Kung Yi Su, Matthew Orr, Denise Schmitz, Ivanna Escala, Robyn Sanderson, Michael Y Grudić, Zachary Hafen, Ji Hoon Kim, Alex Fitts, James S Bullock, Coral Wheeler, T K Chan, Oliver D Elbert, and Desika Narayanan. FIRE-2 simulations: Physics versus numerics in galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 480(1):800–863, 6 2018. ISSN 13652966. doi: 10.1093/mnras/sty1690. URL http://arxiv.org/abs/1702.06148https://academic.oup.com/mnras/advance-article/doi/10.1093/mnras/sty1690/5046474.

[55] Andrew Pontzen, Michael Tremmel, Nina Roth, Hiranya V. Peiris, AmÃĺlie Saintonge, Marta Volonteri, Tom Quinn, and Fabio Governato. How to quench a galaxy. *Monthly Notices of the Royal Astronomical Society*, 465(1):547–558, 2 2017. ISSN 0035-8711. doi: 10.1093/mnras/stw2627. URL https://arxiv.org/pdf/1607.02507.pdfhttps://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw2627.

[56] Simon D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183(3):341–358, 7 1978. ISSN 0035-8711. doi: 10.1093/mnras/183.3.341. URL https://academic.oup.com/mnras/article-abstract/183/3/341/972568https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/183.3.341.

[57] R G Bower, A J Benson, R Malbon, J C Helly, C S Frenk, C M Baugh, S Cole, and C G Lacey. Breaking the hierarchy of galaxy formation. *Monthly Notices of*

the Royal Astronomical Society, 370(2):645–655, 8 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.10519.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/370/2/10.1111/j.1365-2966.2006.10519.x/2/mnras0370-0645.pdf?Expires=1490611456&Signature=Qnt9a2~kcM8OvnwzjYR5cnXYp-I1gOy~8Uyed~mVMhUj2UbayXIA4VmmrRJUewL72lfavq8X-Jd0zn4t~7

[58] Gabriella De Lucia, Fabio Fontanot, David Wilman, and Pierluigi Monaco. Times, environments and channels of bulge formation in a Lambda cold dark matter cosmology. *Monthly Notices of the Royal Astronomical Society*, 414(2):1439–1454, 2011. ISSN 00358711. doi: 10.1111/j.1365-2966.2011.18475.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/414/2/10.1111/j.1365-2966.2011.18475.x/2/mnras0414-1439.pdf?Expires=1496235351&Signature=fldYhQDZ2-CP8zQxtxdBetdRlWbjEOOhy658tw8KX36fZGyjbUfRKeqs0~dmTWzrPYOo2l3JrvSldGY~75

[59] Fabio Fontanot, Gabriella de Lucia, David Wilman, and Pierluigi Monaco. The other side of bulge formation in a Λ cold dark matter cosmology: Bulgeless galaxies in the local Universe. *Monthly Notices of the Royal Astronomical Society*, 416(1):409–415, 2011. ISSN 00358711. doi: 10.1111/j.1365-2966.2011.19047.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/416/1/10.1111/j.1365-2966.2011.19047.x/2/mnras0416-0409.pdf?Expires=1496510229&Signature=TmQAv1RuxjDOI-ZQJMLf1tv9oWIkiF4-VM~1OaK2vt1Qp1l75ahJdrFr3jipRUh9T0r4OXahMaV5oQ3uuY

[60] Rachel S. Somerville, Rudy C. Gilmore, Joel R. Primack, and Alberto Domínguez. Galaxy properties from the ultraviolet to the far-infrared: Λ cold dark matter models confront observations. *Monthly Notices of the Royal Astronomical Society*, 423(3):1992–2015, 2012. ISSN 00358711. doi: 10.1111/j.1365-2966.2012.20490.x.

[61] L. A. Porter, R. S. Somerville, J. R. Primack, D. J. Croton, M. D. Covington, G. J. Graves, and S. M. Faber. Modelling the ages and metallicities of early-type galaxies in Fundamental plane space. *Monthly Notices of the Royal Astronomical Society*, 445(3):3092–3104, 2014. ISSN 13652966. doi: 10.1093/mnras/stu1701.

[62] Simon D M White and Carlos S. Frenk. Galaxy formation through hierarchical clustering. *The Astrophysical Journal*, 379:52, 1991. ISSN 0004-637X. doi: 10.1086/170483.

[63] S Chandrasekhar. DYNAMICAL FRICTION I. GENERAL CONSIDERATIONS: THE COEFFICIENT OF DYNAMICAL FRICTION. *Astrophysical Journal*, 97:255, 1943. URL http://articles.adsabs.harvard.edu/cgi-bin/

nph-iarticle_query?1943ApJ....97..255C&data_type=PDF_HIGH&whole_paper=YES&type=PRINTER&filetype=.pdf.

[64] Michael Boylan-Kolchin, Chung Pei Ma, and Eliot Quataert. Dynamical friction and galaxy merging time-scales. *Monthly Notices of the Royal Astronomical Society*, 383(1):93–101, 2008. ISSN 00358711. doi: 10.1111/j.1365-2966.2007.12530.x.

[65] Philip F Hopkins, Thomas J Cox, Joshua D Younger, and Lars Hernquist. HOW DO DISKS SURVIVE MERGERS? *The Astrophysical Journal*, 691(2): 1168–1201, 2 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/691/2/1168. URL https://arxiv.org/pdf/0806.1739.pdfhttp://stacks.iop.org/0004-637X/691/i=2/a=1168?key=crossref.14f142020ff4890c16f0f2318aed384d.

[66] Simon Richard, Chris B. Brook, Hugo Martel, Daisuke Kawata, Brad K. Gibson, and Patricia Sanchez-Blazquez. Structure, kinematics and chemical enrichment patterns after major gas-rich disc-disc mergers. *Monthly Notices of the Royal Astronomical Society*, 402(3):1489–1503, 11 2010. ISSN 13652966. doi: 10.1111/j.1365-2966.2009.16008.x. URL http://arxiv.org/abs/0911.1801http://dx.doi.org/10.1111/j.1365-2966.2009.16008.x.

[67] Robert C. Kennicutt, Jr. The Global Schmidt Law in StarâĂŘforming Galaxies. *The Astrophysical Journal*, 498(2):541–552, 1998. ISSN 0004-637X. doi: 10.1086/305588.

[68] Frank Bigiel, Adam Leroy, and Fabian Walter. The star formation law at sub-kiloparsec resolution. In *Astrophysics and Space Science Proceedings*, number 208949, pages 105–108. Springer Netherlands, 2008. doi: 10.1007/978-1-4020-6933-8{_}22. URL http://link.springer.com/10.1007/978-1-4020-6933-8_22.

[69] Desika Narayanan, Mark R. Krumholz, Eve C. Ostriker, and Lars Hernquist. A general model for the CO-H 2 conversion factor in galaxies with applications to the star formation law. *Monthly Notices of the Royal Astronomical Society*, 421 (4):3127–3146, 2012. ISSN 00358711. doi: 10.1111/j.1365-2966.2012.20536.x.

[70] Darren Croton, Croton, and Darren. The Environments Of Galaxies As Triggers For AGN Driven Star Formation 'quenching'. *AAS*, 211:125.04, 2007.

[71] Rachel Asquith, Frazer R Pearce, Omar Almaini, Alexander Knebe, Violeta Gonzalez-Perez, Andrew Benson, Jeremy Blaizot, Jorge Carretero, Francisco J Castander, Andrea Cattaneo, SofÃŋa A Cora, Darren J Croton, Julien E Devriendt, Fabio Fontanot, Ignacio D Gargiulo, Will Hartley, Bruno Henriques, Jaehyun Lee, Gary A Mamon, Julian Onions, Nelson D Padilla, Chris Power, Chaichalit

Srisawat, Adam R H Stevens, Peter A Thomas, Cristian A Vega-Martínez, and Sukyoung K Yi. Cosmic CARNage II: the evolution of the galaxy stellar mass function in observations and galaxy formation models. *Monthly Notices of the Royal Astronomical Society*, 480(1):1197–1210, 10 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty1870. URL https://arxiv.org/pdf/1807.03796.pdfhttp://arxiv.org/abs/1807.03796%0Ahttp://dx.doi.org/10.1093/mnras/sty1870https://academic.oup.com/mnras/article/480/1/1197/5054051.

[72] Owain N. Snaith, Brad K. Gibson, Chris B. Brook, StÃľphanie Courty, Patricia Sánchez-Blázquez, Daisuke Kawata, Alexander Knebe, and Laura V. Sales. A comparison of galaxy group luminosity functions from semi-analytic models. *Monthly Notices of the Royal Astronomical Society*, 415(3):2798–2811, 8 2011. ISSN 00358711. doi: 10.1111/j.1365-2966.2011.18907.x.

[73] G. De Lucia and J. Blaizot. The hierarchical formation of the brightest cluster galaxies. *Monthly Notices of the Royal Astronomical Society*, 375(1):2–14, 2 2007. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.11287.x. URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2006.11287.x.

[74] Michaela Hirschmann, Gabriella De Lucia, and Fabio Fontanot. Galaxy assembly, stellar feedback and metal enrichment: The view from the gaea model. *Monthly Notices of the Royal Astronomical Society*, 461(2):1760–1785, 2016. ISSN 13652966. doi: 10.1093/mnras/stw1318. URL http://gavo.mpa-garching.mpg.de/Millennium/.

[75] Lizhi Xie, Gabriella De Lucia, Michaela Hirschmann, Fabio Fontanot, and Anna Zoldan. H2-based star formation laws in hierarchical models of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 469(1):968–993, 7 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx889. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stx889.

[76] Anna Zoldan, Gabriella De Lucia, Lizhi Xie, Fabio Fontanot, and Michaela Hirschmann. The evolution of sizes and specific angular momenta in hierarchical models of galaxy formation and evolution. *Monthly Notices of the Royal Astronomical Society*, 487(4):5649–5665, 8 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1670. URL https://academic.oup.com/mnras/article/487/4/5649/5520834.

[77] Claudia Del P Lagos, Rodrigo J Tobar, Aaron S.G. G Robotham, Danail Obreschkow, Peter D Mitchell, Chris Power, and Pascal J Elahi. Shark: Introducing an open source, free, and flexible semi-analytic model of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 481(3):

3573–3603, 12 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty2440. URL https://github.com/pelahi/VELOCIraptor-STFhttps://academic.oup.com/mnras/article/481/3/3573/5091819.

[78] Claudia Del P. Lagos, Aaron S.G. Robotham, James W. Trayford, Rodrigo Tobar, MatÃŋas Bravo, Sabine Bellstedt, Luke J.M. Davies, Simon P. Driver, Pascal J. Elahi, Danail Obreschkow, and Chris Power. From the far-ultraviolet to the far-infrared - Galaxy emission at 0 âĽď z âĽď 10 in the SHARK semi-analytic model. *Monthly Notices of the Royal Astronomical Society*, 489(3):4196–4216, 2019. ISSN 13652966. doi: 10.1093/mnras/stz2427. URL https://github.com/ICRAR/shark.

[79] Andrea Lapi and Alfonso Cavaliere. Dark Matter Halos: The Dynamical Basis of Effective Empirical Models. *Advances in Astronomy*, 2011:1–8, 10 2011. ISSN 1687-7969. doi: 10.1155/2011/903429. URL https://arxiv.org/pdf/1010.2602.pdfhttp://arxiv.org/abs/1010.2602http://dx.doi.org/10.1155/2011/903429http://www.hindawi.com/journals/aa/2011/903429/.

[80] Valentino González, Ivo Labbé, Rychard J Bouwens, Garth Illingworth, Marijn Franx, and Mariska Kriek. Evolution of galaxy stellar mass functions, mass densities, and mass-to-light ratios from z âĹij 7 to z âĹij 4. *Astrophysical Journal Letters*, 735(2):L34, 7 2011. ISSN 20418205. doi: 10.1088/2041-8205/735/2/L34. URL https://arxiv.org/pdf/1010.2602.pdfhttp://arxiv.org/abs/1010.2602http://dx.doi.org/10.1155/2011/903429http://www.hindawi.com/journals/aa/2011/903429/http://stacks.iop.org/2041-8205/735/i=2/a=L34?key=crossref.e38442dad411fb5d885e672d093d1aea.

[81] Andrey V Kravtsov, Andreas A Berlind, Risa H Wechsler, Anatoly A Klypin, Stefan Gottlober, Brandon Allgood, and Joel R Primack. The Dark Side of the Halo Occupation Distribution. *The Astrophysical Journal*, 609(1):35–49, 7 2004. ISSN 0004-637X. doi: 10.1086/420959. URL http://www-hpcc.astro.washington.edu/tools/tools.html.http://arxiv.org/abs/astro-ph/0308519%0Ahttp://dx.doi.org/10.1086/420959http://arxiv.org/abs/astro-ph/0308519http://dx.doi.org/10.1086/420959http://stacks.iop.org/0004-637X/609/i=1/a=35.

[82] F Shankar, A Lapi, P Salucci, G De Zotti, and L Danese. New Relationships between Galaxy Properties and Host Halo Mass, and the Role of Feedbacks in Galaxy Formation. *The Astrophysical Journal*, 643(1):14–25, 5 2006. ISSN 0004-637X. doi: 10.1086/502794. URL http://iopscience.iop.org/article/10.1086/502794/pdfhttp://adsabs.harvard.edu/abs/2006ApJ...643...14Shttp://stacks.iop.org/0004-637X/643/i=1/a=14.

[83] Joel Leja, Pieter G Van Dokkum, Marijn Franx, and Katherine E Whitaker. Reconciling the observed star-forming sequence with the observed stellar mass function. *Astrophysical Journal*, 798(2):115–16, 2015. ISSN 15384357. doi: 10.1088/0004-637X/798/2/115. URL http://iopscience.iop.org/article/10.1088/0004-637X/798/2/115/pdf.

[84] A. Lapi, Claudia Mancuso, A. Bressan, and L. Danese. Stellar Mass Function of Active and Quiescent Galaxies via the Continuity Equation. *The Astrophysical Journal*, 847(1):13, 9 2017. ISSN 1538-4357. doi: 10.3847/1538-4357/aa88c9. URL http://arxiv.org/abs/1708.07643http://dx.doi.org/10.3847/1538-4357/aa88c9http://stacks.iop.org/0004-637X/847/i=1/a=13?key=crossref.3feb3a1a63828e7e4021664bde43e451.

[85] Aldo Rodríguez-Puebla, Peter Behroozi, Joel Primack, Anatoly Klypin, Christoph Lee, and Doug Hellinger. Halo and subhalo demographics with Planck cosmological parameters: Bolshoi-Planck and MultiDark-Planck simulations. *Monthly Notices of the Royal Astronomical Society*, 462(1):893–916, 2016. ISSN 13652966. doi: 10.1093/mnras/stw1705.

[86] Peter S. Behroozi, Risa H. Wechsler, Hao Yi Wu, Michael T. Busha, Anatoly A. Klypin, and Joel R. Primack. Gravitationally consistent halo catalogs and merger trees for precision cosmology. *Astrophysical Journal*, 763 (1):18, 1 2013. ISSN 15384357. doi: 10.1088/0004-637X/763/1/18. URL http://code.google.com/p/consistent-trees.http://hipacc.ucsc.edu/Bolshoi/.http://stacks.iop.org/0004-637X/763/i=1/a=18?key=crossref.49c9e295a6c0de6f896bce1dbbcecf88.

[87] Romina Ahumada, Carlos Allende Prieto, Andres Almeida, Friedrich Anders, Scott F. Anderson, Brett H. Andrews, Borja Anguiano, Riccardo Arcodia, Eric Armengaud, Marie Aubert, Santiago Avila, Vladimir Avila-Reese, Carles Badenes, Christophe Balland, Kat Barger, Jorge K. Barrera-Ballesteros, Sarbani Basu, Julian Bautista, Rachael L. Beaton, Timothy C. Beers, B. Izamar T. Benavides, Chad F. Bender, Mariangela Bernardi, Matthew Bershady, Florian Beutler, Christian Moni Bidin, Jonathan Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Mederic Boquien, Jura Borissova, Jo Bovy, W. N. Brandt, Jonathan Brinkmann, Joel R. Brownstein, Kevin Bundy, Martin Bureau, Adam Burgasser, Etienne Burtin, Mariana Cano-Diaz, Raffaella Capasso, Michele Cappellari, Ricardo Carrera, Solene Chabanier, William Chaplin, Michael Chapman, Brian Cherinka, Cristina Chiappini, Peter Doohyun Choi, S. Drew Chojnowski, Haeun Chung, Nicolas Clerc, Damien Coffey, Julia M. Comerford, Johan Comparat, Luiz da Costa, Marie-Claude Cousinou, Kevin Covey, Jeffrey D. Crane, Katia

Cunha, Gabriele da Silva Ilha, Yu Sophia Dai, Sanna B. Damsted, Jeremy Darling, Danny Horta Darrington, James W. Davidson, Roger Davies, Kyle Dawson, Nikhil De, Axel de la Macorra, Nathan De Lee, Anna Barbara de Andrade Queiroz, Alice Deconto Machado, Sylvain de la Torre, Flavia Dell'Agli, Helion du Mas des Bourboux, Aleksandar M. Diamond-Stanic, Sean Dillon, John Donor, Niv Drory, Chris Duckworth, Tom Dwelly, Garrett Ebelke, Sarah Eftekharzadeh, Arthur Davis Eigenbrot, Yvonne P. Elsworth, Mike Eracleous, Ghazaleh Erfanianfar, Stephanie Escoffier, Xiaohui Fan, Emily Farr, Jose G. Fernandez-Trincado, Diane Feuillet, Alexis Finoguenov, Patricia Fofie, Amelia Fraser-McKelvie, Peter M. Frinchaboy, Sebastien Fromenteau, Hai Fu, Lluis Galbany, Rafael A. Garcia, D. A. Garcia-Hernandez, Luis Alberto Garma Oehmichen, Junqiang Ge, Marcio Antonio Geimba Maia, Doug Geisler, Joseph Gelfand, Julian Goddy, Jean-Marc Le Goff, Violeta Gonzalez-Perez, Kathleen Grabowski, Paul Green, Catherine J. Grier, Hong Guo, Julien Guy, Paul Harding, Sten Hasselquist, Adam James Hawken, Christian R. Hayes, Fred Hearty, S. Hekker, David W. Hogg, Jon Holtzman, Jiamin Hou, Bau-Ching Hsieh, Daniel Huber, Jason A. S. Hunt, J. Ider Chitham, Julie Imig, Mariana Jaber, Camilo Eduardo Jimenez Angel, Jennifer A. Johnson, Amy M. Jones, Henrik Jonsson, Eric Jullo, Yerim Kim, Karen Kinemuchi, Charles C. Kirkpatrick, George W. Kite, Mark Klaene, Jean-Paul Kneib, Juna A. Kollmeier, Hui Kong, Marina Kounkel, Dhanesh Krishnarao, Ivan Lacerna, Ting-Wen Lan, Richard R. Lane, David R. Law, Henry W. Leung, Hannah Lewis, Cheng Li, Jianhui Lian, Lihwai Lin, Dan Long, Penelope Longa-Pena, Britt Lundgren, Brad W. Lyke, J. Ted Mackereth, Chelsea L. MacLeod, Steven R. Majewski, Arturo Manchado, Claudia Maraston, Paul Martini, Thomas Masseron, Karen L. Masters, Savita Mathur, Richard M. McDermid, Andrea Merloni, Michael Merrifield, Szabolcs Meszaros, Andrea Miglio, Dante Minniti, Rebecca Minsley, Takamitsu Miyaji, Faizan Gohar Mohammad, Benoit Mosser, Eva-Maria Mueller, Demitri Muna, Andrea Munoz-Gutierrez, Adam D. Myers, Seshadri Nadathur, Preethi Nair, Janaina Correa do Nascimento, Rebecca Jean Nevin, Jeffrey A. Newman, David L. Nidever, Christian Nitschelm, Pasquier Noterdaeme, Julia E. O'Connell, Matthew D Olmstead, Daniel Oravetz, Audrey Oravetz, Yeisson Osorio, Zachary J. Pace, Nelson Padilla, Nathalie Palanque-Delabrouille, Pedro A. Palicio, Hsi-An Pan, Kaike Pan, James Parker, Romain Paviot, Sebastien Peirani, Karla Pena Ramrez, Samantha Penny, Will J. Percival, Ismael Perez-Fournon, Ignasi Perez-Rafols, Patrick Petitjean, Matthew M. Pieri, Marc Pinsonneault, Vijith Jacob Poovelil, Joshua Tyler Povick, Abhishek Prakash, Adrian M. Price-Whelan, M. Jordan Raddick, Anand Raichoor, Amy Ray, Sandro Barboza Rembold, Mehdi Rezaie, Rogemar A. Riffel, Rogerio Riffel, Hans-Walter Rix, Annie C. Robin, A. Roman-Lopes, Carlos Roman-Zuniga,

Benjamin Rose, Ashley J. Ross, Graziano Rossi, Kate Rowlands, Kate H. R. Rubin, Mara Salvato, Ariel G. Sanchez, Laura Sanchez-Menguiano, Jose R. Sanchez-Gallego, Conor Sayres, Adam Schaefer, Ricardo P. Schiavon, Jaderson S. Schimoia, Edward Schlafly, David Schlegel, Donald P. Schneider, Mathias Schultheis, Axel Schwope, Hee-Jong Seo, Aldo Serenelli, Arman Shafieloo, Shoaib Jamal Shamsi, Zhengyi Shao, Shiyin Shen, Matthew Shetrone, Raphael Shirley, Victor Silva Aguirre, Joshua D. Simon, M. F. Skrutskie, Anze Slosar, Rebecca Smethurst, Jennifer Sobeck, Bernardo Cervantes Sodi, Diogo Souto, David V. Stark, Keivan G. Stassun, Matthias Steinmetz, Dennis Stello, Julianna Stermer, Thaisa Storchi-Bergmann, Alina Streblyanska, Guy S. Stringfellow, Amelia Stutz, Genaro Suarez, Jing Sun, Manuchehr Taghizadeh-Popp, Michael S. Talbot, Jamie Tayar, Aniruddha R. Thakar, Riley Theriault, Daniel Thomas, Zak C. Thomas, Jeremy Tinker, Rita Tojeiro, Hector Hernandez Toledo, Christy A. Tremonti, Nicholas W. Troup, Sarah Tuttle, Eduardo Unda-Sanzana, Marica Valentini, Jaime Vargas-Gonzalez, Mariana Vargas-Magana, Jose Antonio Vazquez-Mata, M. Vivek, David Wake, Yuting Wang, Benjamin Alan Weaver, Anne-Marie Weijmans, Vivienne Wild, John C. Wilson, Robert F. Wilson, Nathan Wolthuis, W. M. Wood-Vasey, Renbin Yan, Meng Yang, Christophe Yeche, Olga Zamora, Pauline Zarrouk, Gail Zasowski, Kai Zhang, Cheng Zhao, Gongbo Zhao, Zheng Zheng, Zheng Zheng, Guangtun Zhu, and Hu Zou. The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. *Anna Bárbara de Andrade Queiroz*, 43:17, 12 2019. URL http://arxiv.org/abs/1912.02905.

[88] Norman A Grogin, Dale D Kocevski, S M Faber, Henry C Ferguson, Anton M Koekemoer, Adam G Riess, Viviana Acquaviva, David M Alexander, Omar Almaini, Matthew L.N. Ashby, Marco Barden, Eric F Bell, FrÃĭdÃĭric Bournaud, Thomas M Brown, Karina I Caputi, Stefano Casertano, Paolo Cassata, Marco Castellano, Peter Challis, Ranga Ram Chary, Edmond Cheung, Michele Cirasuolo, Christopher J Conselice, Asantha Roshan Cooray, Darren J Croton, Emanuele Daddi, Tomas Dahlen, Romeel Davé, DuÃŋlia F. De Mello, Avishai Dekel, Mark Dickinson, Timothy Dolch, Jennifer L Donley, James S Dunlop, Aaron A Dutton, David Elbaz, Giovanni G Fazio, Alexei V Filippenko, Steven L Finkelstein, Adriano Fontana, Jonathan P Gardner, Peter M Garnavich, Eric Gawiser, Mauro Giavalisco, Andrea Grazian, Yicheng Guo, Nimish P Hathi, Boris Häussler, Philip F Hopkins, Jia Sheng Huang, Kuang Han Huang, Saurabh W Jha, Jeyhan S Kartaltepe, Robert P Kirshner, David C Koo, Kamson Lai, Kyoung Soo Lee, Weidong Li,

Jennifer M Lotz, Ray A Lucas, Piero Madau, Patrick J McCarthy, Elizabeth J McGrath, Daniel H McIntosh, Ross J McLure, Bahram Mobasher, Leonidas A Moustakas, Mark Mozena, Kirpal Nandra, Jeffrey A Newman, Sami Matias Niemi, Kai G Noeske, Casey J Papovich, Laura Pentericci, Alexandra Pope, Joel R Primack, Abhijith Rajan, Swara Ravindranath, Naveen A Reddy, Alvio Renzini, Hans Walter Rix, Aday R Robaina, Steven A Rodney, David J Rosario, Piero Rosati, Sara Salimbeni, Claudia Scarlata, Brian Siana, Luc Simard, Joseph Smidt, Rachel S Somerville, Hyron Spinrad, Amber N Straughn, Louis Gregory Strolger, Olivia Telford, Harry I Teplitz, Jonathan R Trump, Arjen Van Der Wel, Carolin Villforth, Risa H Wechsler, Benjamin J Weiner, Tommy Wiklind, Vivienne Wild, Grant Wilson, Stijn Wuyts, Hao Jing Yan, and Min S Yun. Candels: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey. *Astrophysical Journal, Supplement Series*, 197(2):2011, 2011. ISSN 00670049. doi: 10.1088/0067-0049/197/2/35.

[89] Home Page | COSMOS. URL http://cosmos.astro.caltech.edu/.

[90] Gabriel B. Brammer, Pieter G. Van Dokkum, Marijn Franx, Mattia Fumagalli, Shannon Patel, Hans Walter Rix, Rosalind E. Skelton, Mariska Kriek, Erica Nelson, Kasper B. Schmidt, Rachel Bezanson, Elisabete Da Cunha, Dawn K. Erb, Xiaohui Fan, Natascha Förster Schreiber, Garth D. Illingworth, Ivo Labbé, Joel Leja, Britt Lundgren, Dan Magee, Danilo Marchesini, Patrick McCarthy, Ivelina Momcheva, Adam Muzzin, Ryan Quadri, Charles C. Steidel, Tomer Tal, David Wake, Katherine E. Whitaker, and Anna Williams. 3D-HST: A wide-field grism spectroscopic survey with the hubble space telescope. *Astrophysical Journal, Supplement Series*, 200(2), 2012. ISSN 00670049. doi: 10.1088/0067-0049/200/2/13. URL http://archive.stsci.edu/prepds/hudf09/.

[91] Rogier A Windhorst. JWST White Paper - Galaxies Across Cosmic Time with JWST - Windhorst et al. (2009). Technical report, 2009.

[92] Luca Amendola, Stephen Appleby, Anastasios Avgoustidis, David Bacon, Tessa Baker, Marco Baldi, Nicola Bartolo, Alain Blanchard, Camille Bonvin, Stefano Borgani, Enzo Branchini, Clare Burrage, Stefano Camera, Carmelita Carbone, Luciano Casarini, Mark Cropper, Claudia de Rham, JÃűrg P. Dietrich, Cinzia Di Porto, Ruth Durrer, Anne Ealet, Pedro G. Ferreira, Fabio Finelli, Juan García-Bellido, Tommaso Giannantonio, Luigi Guzzo, Alan Heavens, Lavinia Heisenberg, Catherine Heymans, Henk Hoekstra, Lukas Hollenstein, Rory Holmes, Zhiqi Hwang, Knud Jahnke, Thomas D. Kitching, Tomi Koivisto, Martin Kunz, Giuseppe La Vacca, Eric Linder, Marisa March, Valerio Marra, Carlos Martins, Elisabetta Majerotto, Dida Markovic, David Marsh, Federico Marulli, Richard Massey, Yannick Mellier, Francesco Montanari, David F.

Mota, Nelson J. Nunes, Will Percival, Valeria Pettorino, Cristiano Porciani, Claudia Quercellini, Justin Read, Massimiliano Rinaldi, Domenico Sapone, Ignacy Sawicki, Roberto Scaramella, Constantinos Skordis, Fergus Simpson, Andy Taylor, Shaun Thomas, Roberto Trotta, Licia Verde, Filippo Vernizzi, Adrian Vollmer, Yun Wang, Jochen Weller, and Tom Zlosnik. Cosmology and fundamental physics with the Euclid satellite, 12 2018. ISSN 14338351. URL http://arxiv.org/abs/1606.00180http://dx.doi.org/10.1007/s41114-017-0010-3http://link.springer.com/10.1007/s41114-017-0010-3.

[93] R. Adam, M. Vannier, S. Maurogordato, A. Biviano, C. Adami, B. Ascaso, F. Bellagamba, C. Benoist, A. Cappi, A. Díaz-Sánchez, F. Durret, S. Farrens, A. H. Gonzalez, A. Iovino, R. Licitra, M. Maturi, S. Mei, A. Merson, E. Munari, R. Pelló, M. Ricci, P. F. Rocci, M. Roncarelli, F. Sarron, Y. Amoura, S. Andreon, N. Apostolakos, M. Arnaud, S. Bardelli, J. Bartlett, C. M. Baugh, S. Borgani, M. Brodwin, F. Castander, G. Castignani, O. Cucciati, G. De Lucia, P. Dubath, P. Fosalba, C. Giocoli, H. Hoekstra, G. A. Mamon, J. B. Melin, L. Moscardini, S. Paltani, M. Radovich, B. Sartoris, M. Schultheis, M. Sereno, J. Weller, C. Burigana, C. S. Carvalho, L. Corcione, H. Kurki-Suonio, P. B. Lilje, G. Sirri, R. Toledo-Moreo, and G. Zamorani. Euclid preparation: III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection. *Astronomy and Astrophysics*, 627:38, 6 2019. ISSN 14320746. doi: 10.1051/0004-6361/201935088. URL http://arxiv.org/abs/1906.04707http://dx.doi.org/10.1051/0004-6361/201935088.

[94] A Cattaneo, G A Mamon, K Warnick, and A Knebe. How do galaxies acquire their mass? *A&A*, 533(A5):5, 2011. ISSN 0004-6361. doi: 10.1051/0004-6361/201015780. URL http://popia.ft.uam.es/AMIGA.http://arxiv.org/abs/1002.3257%0Ahttp://dx.doi.org/10.1051/0004-6361/201015780.

[95] Mariangela Bernardi, Nathan Roche, Francesco Shankar, and Ravi K. Sheth. Curvature in the colour-magnitude relation but not in colour-σ: Major dry mergers at M* $>$ 2 ÃŮ 1011MâŁŹ? *Monthly Notices of the Royal Astronomical Society*, 412(1):684–704, 3 2011. ISSN 00358711. doi: 10.1111/j.1365-2966.2010.17984.x. URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2010.17984.x.

[96] Francesco Shankar, Stewart Buchan, Alessandro Rettura, Vincent R Bouillot, Jorge Moreno, Rossella Licitra, Mariangela Bernardi, Marc Huertas-Company, Simona Mei, BegoÃśa Ascaso, Ravi Sheth, Lauriane Delaye, and Anand Raichoor. AVOIDING PROGENITOR BIAS: THE STRUCTURAL AND MASS EVOLUTION OF

BRIGHTEST GROUP AND CLUSTER GALAXIES IN HIERARCHICAL MODELS SINCE z âĽš 1. *The Astrophysical Journal*, 802(2):73, 3 2015. ISSN 1538-4357. doi: 10.1088/0004-637X/802/2/73. URL http://stacks.iop.org/0004-637X/802/i=2/a=73?key=crossref.91ba6eef0f49d4aff7bfd0af72576d83.

[97] Allison W S Man, Andrew W Zirm, and Sune Toft. RESOLVING THE DISCREPANCY OF GALAXY MERGER FRACTION MEASUREMENTS AT z âĹij 0âĂŞ3. *The Astrophysical Journal*, 830(2):89, 10 2016. ISSN 1538-4357. doi: 10.3847/0004-637X/830/2/89. URL http://home.strw.leidenuniv.nl/~ent/InterRest/InterRest-distribution/http://stacks.iop.org/0004-637X/830/i=2/a=89?key=crossref.e2792ce9b809e65a41e2f902b5f3492b.

[98] Philip F Hopkins, Joshua D Younger, Christopher C Hayward, Desika Narayanan, and Lars Hernquist. Mergers, active galactic nuclei and âĂŸnormalâĂŹ galaxies: contributions to the distribution of star formation rates and infrared luminosity functions. *Monthly Notices of the Royal Astronomical Society*, 402 (3):1693–1713, 3 2010. ISSN 00358711. doi: 10.1111/j.1365-2966.2009.15990.x. URL https://arxiv.org/pdf/0911.1131.pdfhttps://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2009.15990.x.

[99] J. Fensch, F. Renaud, F. Bournaud, P.-A. Duc, O. Agertz, P. Amram, F. Combes, P. Di Matteo, B. Elmegreen, E. Emsellem, C. J. Jog, V. Perret, C. Struck, and R. Teyssier. High-redshift major mergers weakly enhance star formation. *Monthly Notices of the Royal Astronomical Society*, 465(2):1934–1949, 2 2017. ISSN 0035-8711. doi: 10.1093/mnras/stw2920. URL http://arxiv.org/abs/1610.03877https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw2920.

[100] G. Martin, S. Kaviraj, J. E.G. Devriendt, Y. Dubois, C. Laigle, and C. Pichon. The limited role of galaxy mergers in driving stellar mass growth over cosmic time. *Monthly Notices of the Royal Astronomical Society: Letters*, 472(1):L50–L54, 2017. ISSN 17453933. doi: 10.1093/mnrasl/slx136. URL https://arxiv.org/pdf/1708.09396.pdf.

[101] FrÃĺdÃĺric Bournaud, Damien Chapon, Romain Teyssier, Leila C. Powell, Bruce G. Elmegreen, Debra Meloy Elmegreen, Pierre Alain Duc, Thierry Contini, Benoit Epinat, and Kristen L. Shapiro. Hydrodynamics of high-redshift galaxy collisions: From gas-rich disks to dispersion-dominated mergers and compact spheroids. *Astrophysical Journal*, 730(1), 2011. ISSN 15384357. doi: 10.1088/0004-637X/730/1/4. URL http://iopscience.iop.org/article/10.1088/0004-637X/730/1/4/pdf.

[102] F. Bournaud. Star formation in galaxy interactions and mergers. *EAS Publications Series*, 51:107–131, 2011. ISSN 1633-4760. doi: 10.1051/eas/1151008. URL https://arxiv.org/pdf/1106.1793.pdf.

[103] Michael Hilz, Thorsten Naab, and J. P. Ostriker. How do minor mergers promote inside-out growth of ellipticals, transforming the size, density profile and dark matter fraction? *Monthly Notices of the Royal Astronomical Society*, 429(4):2924–2933, 2013. ISSN 00358711. doi: 10.1093/mnras/sts501.

[104] Frank C. van den Bosch, Fangzhou Jiang, Andrew Hearin, Duncan Campbell, Douglas Watson, and Nikhil Padmanabhan. Coming of Age in the Dark Sector: How Dark Matter Haloes grow their Gravitational Potential Wells. *MNRAS*, 445 (2):1–20, 9 2014. ISSN 1365-2966. doi: 10.1093/mnras/stu1872.

[105] Fangzhou Jiang and Frank C. van den Bosch. Statistics of dark matter substructure âĂŞ I. Model and universal fitting functions. *Monthly Notices of the Royal Astronomical Society*, 458(3):2848–2869, 5 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw439.

[106] Alexander Knebe, Frazer R. Pearce, Peter A. Thomas, Andrew Benson, Jeremy Blaizot, Richard Bower, Jorge Carretero, Francisco J. Castander, Andrea Cattaneo, Sofia A. Cora, Darren J. Croton, Weiguang Cui, Daniel Cunnama, Gabriella De Lucia, Julien E. Devriendt, Pascal J. Elahi, Andreea Font, Fabio Fontanot, Juan Garcia-Bellido, Ignacio D. Gargiulo, Violeta Gonzalez-Perez, John Helly, Bruno Henriques, Michaela Hirschmann, Jaehyun Lee, Gary A. Mamon, Pierluigi Monaco, Julian Onions, Nelson D. Padilla, Chris Power, Arnau Pujol, Ramin A. Skibba, Rachel S. Somerville, Chaichalit Srisawat, Cristian A. Vega-Martínez, and Sukyoung K. Yi. nIFTy cosmology: comparison of galaxy formation models. *Monthly Notices of the Royal Astronomical Society*, 451(4):4029–4059, 8 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv1149. URL http://popia.ft.uam.es/SussingMergerTreeshttps://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stv1149.

[107] Weiguang Cui, Alexander Knebe, Gustavo Yepes, Frazer Pearce, Chris Power, Romeel Dave, Alexander Arth, Stefano Borgani, Klaus Dolag, Pascal Elahi, Robert Mostoghiu, Giuseppe Murante, Elena Rasia, Doris Stoppacher, Jesus Vega-Ferrero, Yang Wang, Xiaohu Yang, Andrew Benson, SofÃŋa A. Cora, Darren J. Croton, Manodeep Sinha, Adam R.H. Stevens, Cristian A. Vega-Martínez, Jake Arthur, Anna S. Baldi, Rodrigo Cañas, Giammarco Cialone, Daniel Cunnama, Marco De Petris, Giacomo Durando, Stefano Ettori, Stefan Gottlöber, SebastiÃąn E. Nuza, Lyndsay J. Old, Sergey Pilipenko, Jenny G. Sorce, and Charlotte Welker.

The Three Hundred project: A large catalogue of theoretically modelled galaxy clusters for cosmological and astrophysical applications. Technical Report 3, 2018. URL www.cosmosim.org.

[108] Alexander Knebe, Frazer R Pearce, Violeta Gonzalez-Perez, Peter A Thomas, Andrew Benson, Rachel Asquith, Jeremy Blaizot, Richard Bower, Jorge Carretero, Francisco J Castander, Andrea Cattaneo, SofÃŋa A Cora, Darren J Croton, Weiguang Cui, Daniel Cunnama, Julien E Devriendt, Pascal J Elahi, Andreea Font, Fabio Fontanot, Ignacio D Gargiulo, John Helly, Bruno Henriques, Jaehyun Lee, Gary A Mamon, Julian Onions, Nelson D Padilla, Chris Power, Arnau Pujol, AndrÃĺs N Ruiz, Chaichalit Srisawat, Adam R H Stevens, Edouard Tollet, Cristian A Vega-Martínez, and Sukyoung K Yi. Cosmic CARNage I: on the calibration of galaxy formation models. *Monthly Notices of the Royal Astronomical Society*, 475(3):2936–2954, 4 2018. ISSN 0035-8711. doi: 10.1093/mnras/stx3274. URL https://cosmohub.pic.eshttps://academic.oup.com/mnras/article/475/3/2936/4768290.

[109] Martin P. Rey and Andrew Pontzen. Quadratic genetic modifications: A streamlined route to cosmological simulations with controlled merger history. *Monthly Notices of the Royal Astronomical Society*, 474(1):45–54, 6 2018. ISSN 13652966. doi: 10.1093/mnras/stx2744. URL http://arxiv.org/abs/1706.04615http://dx.doi.org/10.1093/mnras/stx2744.

[110] Dylan Nelson, Annalisa Pillepich, Shy Genel, Mark Vogelsberger, Volker Springel, Paul Torrey, Vicente Rodriguez-Gomez, Debora Sijacki, G.F. Snyder, Brendan Griffen, Federico Marinacci, Laura Blecha, Laura Sales, Dandan Xu, and Lars Hernquist. The illustris simulation: Public data release. *Astronomy and Computing*, 13:12–37, 11 2015. ISSN 22131337. doi: 10.1016/j.ascom.2015.09.003. URL https://arxiv.org/pdf/1504.00362.pdfhttps://linkinghub.elsevier.com/retrieve/pii/S2213133715000864.

[111] Frank C. van den Bosch. Dissecting the evolution of dark matter subhaloes in the bolshoi simulation. *Monthly Notices of the Royal Astronomical Society*, 468(1):885–909, 2017. ISSN 13652966. doi: 10.1093/mnras/stx520.

[112] Frank C. van den Bosch, Go Ogiya, Oliver Hahn, and Andreas Burkert. Disruption of Dark Matter Substructure: Fact or Fiction? *MNRAS*, 474(3):3043–3066, 2018. ISSN 0035-8711. doi: 10.1093/mnras/stx2956. URL http://arxiv.org/abs/1711.05276.

[113] Giulia Despali, Carlo Giocoli, Raul E Angulo, Giuseppe Tormen, Ravi K Sheth, Giacomo Baso, and Lauro Moscardini. The universality of the virial halo mass

function and models for non-universality of other halo definitions. *Monthly Notices of the Royal Astronomical Society*, 456(3):2486–2504, 2016. ISSN 13652966. doi: 10.1093/mnras/stv2842. URL http://www.

[114] Benedikt Diemer. COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos. Technical Report 2, Harvard-Smithsonian Center for Astrophysics, 12 2018. URL http://arxiv.org/abs/1712.04512.

[115] Stuart P.D. Gill, Alexander Knebe, Brad K. Gibson, and Michael A. Dopita. The evolution of substructure - II. Linking dynamics to environment. *Monthly Notices of the Royal Astronomical Society*, 351(2):410–422, 6 2004. ISSN 00358711. doi: 10.1111/j.1365-2966.2004.07913.x. URL http://arxiv.org/abs/astro-ph/0404255http://dx.doi.org/10.1111/j.1365-2966.2004.07913.x.

[116] Alan Meert, Vinu Vikram, and Mariangela Bernardi. A catalogue of 2D photometric decompositions in the SDSS-DR7 spectroscopic main galaxy sample: preferred models and systematics. *MNRAS*, 446(4):3943–3974, 2015. doi: 10.1093/mnras/stu2333. URL https://watermark.silverchair.com/stu2333.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAcAwggG8BgkqhkiG9w0BBwagggGtMIIBqQIBADCCAaIGCSqGSIb3DQEHATAeBglgh3h.

[117] Kevork N Abazajian, Jennifer K. Adelman-McCarthy, Marcel A. Agüeros, Sahar S Allam, Carlos Allende Prieto, Deokkeun An, Kurt S. J. Anderson, Scott F Anderson, James Annis, Neta A Bahcall, C A L Bailer-Jones, J C Barentine, Bruce A Bassett, Andrew C Becker, Timothy C Beers, Eric F Bell, Vasily Belokurov, Andreas A Berlind, Eileen F Berman, Mariangela Bernardi, Steven J Bickerton, Dmitry Bizyaev, John P Blakeslee, Michael R Blanton, John J Bochanski, William N Boroski, Howard J Brewington, Jarle Brinchmann, J Brinkmann, Robert J Brunner, TamÃąs Budavári, Larry N Carey, Samuel Carliles, Michael A Carr, Francisco J Castander, David Cinabro, A J Connolly, IstvÃąn Csabai, Carlos E. Cunha, Paul C. Czarapata, James R. A. Davenport, Ernst de Haas, Ben Dilday, Mamoru Doi, Daniel J Eisenstein, Michael L. Evans, N. W. Evans, Xiaohui Fan, Scott D. Friedman, Joshua A. Frieman, Masataka Fukugita, Boris T. Gänsicke, Evalyn Gates, Bruce Gillespie, G. Gilmore, Belinda Gonzalez, Carlos F Gonzalez, Eva K Grebel, James E Gunn, Zsuzsanna Györy, Patrick B Hall, Paul Harding, Frederick H Harris, Michael Harvanek, Suzanne L Hawley, Jeffrey J E Hayes, Timothy M Heckman, John S Hendry, Gregory S Hennessy, Robert B Hindsley, J Hoblitt, Craig J Hogan, David W Hogg, Jon A Holtzman, Joseph B Hyde, Shin-ichi Ichikawa, Takashi Ichikawa, Myungshin Im, Å¡eljko Ivezić, Sebastian Jester, Linhua Jiang, Jennifer A Johnson, Anders M Jorgensen, Mario Jurić,

Stephen M Kent, R Kessler, S J Kleinman, G R Knapp, Kohki Konishi, Richard G Kron, Jurek Krzesinski, Nikolay Kuropatkin, Hubert Lampeitl, Svetlana Lebedeva, Myung Gyoon Lee, Young Sun Lee, R French Leger, SÃľbastien Lépine, Nolan Li, Marcos Lima, Huan Lin, Daniel C Long, Craig P Loomis, Jon Loveday, Robert H Lupton, Eugene Magnier, Olena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Bruce Margon, John P Marriner, David Martínez-Delgado, Takahiko Matsubara, Peregrine M. McGehee, Timothy A. McKay, Avery Meiksin, Heather L Morrison, Fergal Mullally, Jeffrey A Munn, Tara Murphy, Thomas Nash, Ada Nebot, Eric H Neilsen, Heidi Jo Newberg, Peter R Newman, Robert C Nichol, Tom Nicinski, Maria Nieto-Santisteban, Atsuko Nitta, Sadanori Okamura, Daniel J Oravetz, Jeremiah P Ostriker, Russell Owen, Nikhil Padmanabhan, Kaike Pan, Changbom Park, George Pauls, John Peoples, Will J. Percival, Jeffrey R. Pier, Adrian C. Pope, Dimitri Pourbaix, Paul A. Price, Norbert Purger, Thomas Quinn, M. Jordan Raddick, Paola Re Fiorentin, Gordon T. Richards, Michael W Richmond, Adam G Riess, Hans-Walter Rix, Constance M Rockosi, Masao Sako, David J Schlegel, Donald P Schneider, Ralf-Dieter Scholz, Matthias R Schreiber, Axel D Schwope, UroÅą Seljak, Branimir Sesar, Erin Sheldon, Kazu Shimasaku, Valena C Sibley, A E Simmons, Thirupathi Sivarani, J. Allyn Smith, Martin C Smith, Vernesa Smolčić, Stephanie A Snedden, Albert Stebbins, Matthias Steinmetz, Chris Stoughton, Michael A Strauss, Mark SubbaRao, Yasushi Suto, Alexander S. Szalay, IstvÃąn Szapudi, Paula Szkody, Masayuki Tanaka, Max Tegmark, Luis F A Teodoro, Aniruddha R Thakar, Christy A Tremonti, Douglas L Tucker, Alan Uomoto, Daniel E. Vanden Berk, Jan Vandenberg, S Vidrih, Michael S Vogeley, Wolfgang Voges, Nicole P Vogt, Yogesh Wadadekar, Shannon Watters, David H Weinberg, Andrew A West, Simon D M White, Brian C Wilhite, Alainna C Wonders, Brian Yanny, D R Yocum, Donald G York, Idit Zehavi, Stefano Zibetti, and Daniel B Zucker. THE SEVENTH DATA RELEASE OF THE SLOAN DIGITAL SKY SURVEY. *The Astrophysical Journal Supplement Series*, 182(2):543–558, 6 2009. ISSN 0067-0049. doi: 10.1088/0067-0049/182/2/543.

[118] Xiaohu Yang, H J Mo, Frank C. van den Bosch, Youcai Zhang, and Jiaxin Han. EVOLUTION OF THE GALAXYâĂŞDARK MATTER CONNECTION AND THE ASSEMBLY OF GALAXIES IN DARK MATTER HALOS. *The Astrophysical Journal*, 752(1):41, 6 2012. ISSN 0004-637X. doi: 10.1088/0004-637X/752/1/41. URL http://iopscience.iop.org/article/10.1088/0004-637X/752/1/41/pdfhttp://stacks.iop.org/0004-637X/752/i=1/a=41?key=crossref.df76cb8d717082c084d0ec7787d2d033.

[119] M Bernardi, A Meert, R K Sheth, M Huertas-Company, C Maraston, F Shankar, and V Vikram. The massive end of the luminosity and stellar mass functions

and clustering from CMASS to SDSS: Evidence for and against passive evolution. *Monthly Notices of the Royal Astronomical Society*, 455(4):4122–4135, 2016. ISSN 13652966. doi: 10.1093/mnras/stv2487. URL https://academic.oup.com/mnras/article-abstract/455/4/4122/1261452.

[120] M. Bernardi, J. L. Fischer, R. K. Sheth, A. Meert, M. Huertas-Company, F. Shankar, and V. Vikram. Comparing PyMorph and SDSS photometry. II. The differences are more than semantics and are not dominated by intracluster light. *MNRAS*, 468(3):2569–2581, 2 2017. URL http://arxiv.org/abs/1702.08527.

[121] Gerard de Vaucouleurs. Recherches sur les Nebuleuses Extragalactiques. *Annales d'Astrophysique*, 11, 1948. URL http://adsabs.harvard.edu/abs/1948AnAp...11..247D.

[122] M Bernardi, A. Meert, R K Sheth, V Vikram, M Huertas-Company, S Mei, and F Shankar. The massive end of the luminosity and stellar mass functions: Dependence on the fit to the light profile. *Monthly Notices of the Royal Astronomical Society*, 436(1):697–704, 11 2013. ISSN 00358711. doi: 10.1093/mnras/stt1607. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stt1607.

[123] I. Davidzon, O. Ilbert, C. Laigle, J. Coupon, H. J. McCracken, I. Delvecchio, D. Masters, P. L. Capak, B. C. Hsieh, L. Tresse, O. Le Fevre, M. Bethermin, Y. Y. Chang, A. L. Faisst, E. Le Floc'h, C. Steinhardt, S. Toft, H. Aussel, C. Dubois, G. Hasinger, M. Salvato, D. B. Sanders, N. Scoville, and J. D. Silverman. The COSMOS2015 galaxy stellar mass function: 13 billion years of stellar mass assembly in 10 snapshots. *Astronomy & Astrophysics*, 605:A70, 1 2017. ISSN 0004-6361. doi: 10.1051/0004-6361/201730419. URL https://arxiv.org/abs/1701.02734http://arxiv.org/abs/1701.02734%0Ahttp://dx.doi.org/10.1051/0004-6361/201730419http://arxiv.org/abs/1701.02734http://dx.doi.org/10.1051/0004-6361/201730419http://www.aanda.org/10.1051/0004-6361/201730419.

[124] G Bruzual and S Charlot. Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344(4):1000–1028, 2003. ISSN 00358711. doi: 10.1046/j.1365-8711.2003.06897.x. URL https://arxiv.org/pdf/astro-ph/0309134.pdf.

[125] J Trevor Mendel, Luc Simard, Michael Palmer, Sara L Ellison, and David R Patton. A CATALOG OF BULGE, DISK, AND TOTAL STELLAR MASS ESTIMATES FOR THE SLOAN DIGITAL SKY SURVEY. *The Astrophysical Journal Supplement Series*, 210(1):3, 12 2014. ISSN 0067-0049. doi: 10.1088/0067-0049/210/1/

3. URL http://people.ucsc.edu/http://stacks.iop.org/0067-0049/210/i=1/a=3?key=crossref.998c68594ba012ec55be315ec9ebab16.

[126] Eric F Bell, Daniel H. McIntosh, Neal Katz, and Martin D Weinberg. The Optical and NearâĂŘInfrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. *The Astrophysical Journal Supplement Series*, 149(2):289–312, 12 2003. ISSN 0067-0049. doi: 10.1086/378847. URL http://www.ipac.caltech.edu/2mass/releases/second/doc/http://stacks.iop.org/0067-0049/149/i=2/a=289.

[127] Edwin E. Salpeter. The Luminosity Function and Stellar Evolution. *The Astrophysical Journal*, 121:161, 1955. ISSN 0004-637X. doi: 10.1086/145971. URL http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1955ApJ...121..161S&data_type=PDF_HIGH&whole_paper=YES&type=PRINTER&filetype=.pdfhttp://adsabs.harvard.edu/doi/10.1086/145971.

[128] Gilles Chabrier. Galactic Stellar and Substellar Initial Mass Function. *Publications of the Astronomical Society of the Pacific*, 115(809):763–795, 7 2003. ISSN 0004-6280. doi: 10.1086/376392. URL http://iopscience.iop.org/article/10.1086/376392/pdfhttp://iopscience.iop.org/article/10.1086/376392.

[129] Bela Abolfathi, D. S. Aguado, Gabriela Aguilar, Carlos Allende Prieto, Andres Almeida, Tonima Tasnim Ananna, Friedrich Anders, Scott F. Anderson, Brett H. Andrews, Borja Anguiano, Alfonso Aragon-Salamanca, Maria Argudo-Fernandez, Eric Armengaud, Metin Ata, Eric Aubourg, Vladimir Avila-Reese, Carles Badenes, Stephen Bailey, Christophe Balland, Kathleen A. Barger, Jorge Barrera-Ballesteros, Curtis Bartosz, Fabienne Bastien, Dominic Bates, Falk Baumgarten, Julian Bautista, Rachael Beaton, Timothy C. Beers, Francesco Belfiore, Chad F. Bender, Mariangela Bernardi, Matthew A. Bershady, Florian Beutler, Jonathan C. Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Mederic Boquien, Jura Borissova, Jo Bovy, Christian Andres Bradna Diaz, William Nielsen Brandt, Jonathan Brinkmann, Joel R. Brownstein, Kevin Bundy, Adam J. Burgasser, Etienne Burtin, Nicolas G. Busca, Caleb I. Canas, Mariana Cano-Diaz, Michele Cappellari, Ricardo Carrera, Andrew R. Casey, Bernardo Cervantes Sodi, Yanping Chen, Brian Cherinka, Cristina Chiappini, Peter Doohyun Choi, Drew Chojnowski, Chia-Hsun Chuang, Haeun Chung, Nicolas Clerc, Roger E. Cohen, Julia M. Comerford, Johan Comparat, Janaina Correa do Nascimento, Luiz da Costa, Marie-Claude Cousinou, Kevin Covey, Jeffrey D. Crane, Irene Cruz-Gonzalez, Katia Cunha, Gabriele da Silva Ilha, Guillermo J. Damke, Jeremy Darling, James W. Davidson, Kyle Dawson, Miguel Angel C. de Icaza Lizaola, Axel de la Macorra, Sylvain de la

Torre, Nathan De Lee, Victoria de Sainte Agathe, Alice Deconto Machado, Flavia Dell'Agli, Timothee Delubac, Aleksandar M. Diamond-Stanic, John Donor, Juan Jose Downes, Niv Drory, Helion du Mas des Bourboux, Christopher J. Duckworth, Tom Dwelly, Jamie Dyer, Garrett Ebelke, Arthur Davis Eigenbrot, Daniel J. Eisenstein, Yvonne P. Elsworth, Eric Emsellem, Mike Eracleous, Ghazaleh Erfanianfar, Stephanie Escoffier, Xiaohui Fan, Emma Fernandez Alvar, J. G. Fernandez-Trincado, Rafael Fernando Cirolini, Diane Feuillet, Alexis Finoguenov, Scott W. Fleming, Andreu Font-Ribera, Gordon Freischlad, Peter Frinchaboy, Hai Fu, Yilen Gomez Maqueo Chew, Lluis Galbany, Ana E. Garcia Perez, R. Garcia-Dias, D. A. Garcia-Hernandez, Luis Alberto Garma Oehmichen, Patrick Gaulme, Joseph Gelfand, Hector Gil-Marin, Bruce A. Gillespie, Daniel Goddard, Jonay I. Gonzalez Hernandez, Violeta Gonzalez-Perez, Kathleen Grabowski, Paul J. Green, Catherine J. Grier, Alain Gueguen, Hong Guo, Julien Guy, Alex Hagen, Patrick Hall, Paul Harding, Sten Hasselquist, Suzanne Hawley, Christian R. Hayes, Fred Hearty, Saskia Hekker, Jesus Hernandez, Hector Hernandez Toledo, David W. Hogg, Kelly Holley-Bockelmann, Jon Holtzman, Jiamin Hou, Bau-Ching Hsieh, Jason A. S. Hunt, Timothy A. Hutchinson, Ho Seong Hwang, Camilo Eduardo Jimenez Angel, Jennifer A. Johnson, Amy Jones, Henrik Jonsson, Eric Jullo, Fahim Sakil Khan, Karen Kinemuchi, David Kirkby, Charles C. Kirkpatrick, Francisco-Shu Kitaura, Gillian R. Knapp, Jean-Paul Kneib, Juna A. Kollmeier, Ivan Lacerna, Richard R. Lane, Dustin Lang, David R. Law, Jean-Marc Le Goff, Young-Bae Lee, Hongyu Li, Cheng Li, Jianhui Lian, Yu Liang, Marcos Lima, Lihwai Lin, Dan Long, Sara Lucatello, Britt Lundgren, J. Ted Mackereth, Chelsea L. MacLeod, Suvrath Mahadevan, Marcio Antonio Geimba Maia, Steven Majewski, Arturo Manchado, Claudia Maraston, Vivek Mariappan, Rui Marques-Chaves, Thomas Masseron, Karen L. Masters, Richard M. McDermid, Ian D. McGreer, Matthew Melendez, Sofia Meneses-Goytia, Andrea Merloni, Michael R. Merrifield, Szabolcs Meszaros, Andres Meza, Ivan Minchev, Dante Minniti, Eva-Maria Mueller, Francisco Muller-Sanchez, Demitri Muna, Ricardo R. Munoz, Adam D. Myers, Preethi Nair, Kirpal Nandra, Melissa Ness, Jeffrey A. Newman, Robert C. Nichol, David L. Nidever, Christian Nitschelm, Pasquier Noterdaeme, Julia O'Connell, Ryan James Oelkers, Audrey Oravetz, Daniel Oravetz, Erik Aquino Ortiz, Yeisson Osorio, Zach Pace, Nelson Padilla, Nathalie Palanque-Delabrouille, Pedro Alonso Palicio, Hsi-An Pan, Kaike Pan, Taniya Parikh, Isabelle Paris, Changbom Park, Sebastien Peirani, Marcos Pellejero-Ibanez, Samantha Penny, Will J. Percival, Ismael Perez-Fournon, Patrick Petitjean, Matthew M. Pieri, Marc Pinsonneault, Alice Pisani, Francisco Prada, Abhishek Prakash, Anna Barbara de Andrade Queiroz,

M. Jordan Raddick, Anand Raichoor, Sandro Barboza Rembold, Hannah Richstein, Rogemar A. Riffel, Rogerio Riffel, Hans-Walter Rix, Annie C. Robin, Sergio Rodriguez Torres, Carlos Roman-Zuniga, Ashley J. Ross, Graziano Rossi, John Ruan, Rossana Ruggeri, Jose Ruiz, Mara Salvato, Ariel G. Sanchez, Sebastian F. Sanchez, Jorge Sanchez Almeida, Jose R. Sanchez-Gallego, Felipe Antonio Santana Rojas, Basilio Xavier Santiago, Ricardo P. Schiavon, Jaderson S. Schimoia, Edward Schlafly, David Schlegel, Donald P. Schneider, William J. Schuster, Axel Schwope, Hee-Jong Seo, Aldo Serenelli, Shiyin Shen, Yue Shen, Matthew Shetrone, Michael Shull, Victor Silva Aguirre, Joshua D. Simon, Mike Skrutskie, Anze Slosar, Rebecca Smethurst, Verne Smith, Jennifer Sobeck, Garrett Somers, Barbara J. Souter, Diogo Souto, Ashley Spindler, David V. Stark, Keivan Stassun, Matthias Steinmetz, Dennis Stello, Thaisa Storchi-Bergmann, Alina Streblyanska, Guy Stringfellow, Genaro Suarez, Jing Sun, Laszlo Szigeti, Manuchehr Taghizadeh-Popp, Michael S. Talbot, Baitian Tang, Charling Tao, Jamie Tayar, Mita Tembe, Johanna Teske, Aniruddha R. Thaker, Daniel Thomas, Patricia Tissera, Rita Tojeiro, Christy Tremonti, Nicholas W. Troup, Meg Urry, O. Valenzuela, Remco van den Bosch, Jaime Vargas-Gonzalez, Mariana Vargas-Magana, Jose Alberto Vazquez, Sandro Villanova, Nicole Vogt, David Wake, Yuting Wang, Benjamin Alan Weaver, Anne-Marie Weijmans, David H. Weinberg, Kyle B. Westfall, David G. Whelan, Eric Wilcots, Vivienne Wild, Rob A. Williams, John Wilson, W. M. Wood-Vasey, Dominika Wylezalek, Ting Xiao, Renbin Yan, Meng Yang, Jason E. Ybarra, Christophe Yeche, Nadia Zakamska, Olga Zamora, Pauline Zarrouk, Gail Zasowski, Kai Zhang, Cheng Zhao, Gong-Bo Zhao, Zheng Zheng, Zheng Zheng, Zhi-Min Zhou, Guangtun Zhu, Joel C. Zinn, and Hu Zou. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the extended Baryon Oscillation Spectroscopic Survey and from the second phase of the Apache Point Observatory Galactic Evolution Experiment. *The Astrophysical Journal Supplement Series*, 235:42, 2017. ISSN 1538-4365. doi: 10.3847/1538-4365/aa9e8a. URL https://doi.org/10.3847/1538-4365/aa9e8ahttp://arxiv.org/abs/1707.09322%0Ahttp://dx.doi.org/10.3847/1538-4365/aa9e8a.

[130] Edward L. Wright, Peter R.M. Eisenhardt, Amy K. Mainzer, Michael E. Ressler, Roc M. Cutri, Thomas Jarrett, J. Davy Kirkpatrick, Deborah Padgett, Robert S. McMillan, Michael Skrutskie, S. A. Stanford, Martin Cohen, Russell G. Walker, John C. Mather, David Leisawitz, Thomas N. Gautier, Ian McLean, Dominic Benford, Carol J. Lonsdale, Andrew Blain, Bryan Mendez, William R. Irace, Valerie Duval, Fengchuan Liu, Don Royer, Ingolf Heinrichsen, Joan Howard, Mark Shannon, Martha Kendall, Amy L. Walsh, Mark Larsen, Joel G. Cardon, Scott Schick, Mark Schwalm, Mohamed Abid, Beth Fabinsky, Larry Naes,

and Chao Wei Tsai. The Wide-field Infrared Survey Explorer (wise): Mission description and initial on-orbit performance. *Astronomical Journal*, 140(6): 1868–1881, 2010. ISSN 00046256. doi: 10.1088/0004-6256/140/6/1868. URL http://iopscience.iop.org/article/10.1088/0004-6256/140/6/1868/pdf.

[131] Benjamin P Moster, Rachel S Somerville, Christian Maulbetsch, Frank C Van Den Bosch, Andrea V Maccì O, Thorsten Naab, and Ludwig Oser. CONSTRAINTS ON THE RELATIONSHIP BETWEEN STELLAR MASS AND HALO MASS AT LOW AND HIGH REDSHIFT. *The Astrophysical Journal*, 710:903–923, 2010. doi: 10.1088/0004-637X/710/2/903.

[132] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee : The MCMC Hammer . *Publications of the Astronomical Society of the Pacific*, 125(925):306–312, 3 2013. ISSN 00046280. doi: 10.1086/670067. URL http://dan.iel.fm/emceeundertheMITLicense.http://iopscience.iop.org/article/10.1086/670067.

[133] Francesco Shankar, Hong Guo, Vincent Bouillot, Alessandro Rettura, Alan Meert, Stewart Buchan, Andrey Kravtsov, Mariangela Bernardi, Ravi Sheth, Vinu Vikram, Danilo Marchesini, Peter Behroozi, Zheng Zheng, Claudia Maraston, BegoÃśa Ascaso, Brian C. Lemaux, Diego Capozzi, Marc Huertas-Company, Roy R. Gal, Lori M. Lubin, Christopher J. Conselice, Marcella Carollo, and Andrea Cattaneo. ON THE INTERMEDIATE-REDSHIFT CENTRAL STELLAR MASS-HALO MASS RELATION, AND IMPLICATIONS FOR THE EVOLUTION OF THE MOST MASSIVE GALAXIES SINCE z âĹij 1. *The Astrophysical Journal*, 797(2):L27, 12 2014. ISSN 2041-8213. doi: 10.1088/2041-8205/797/2/L27. URL http://stacks.iop.org/2041-8205/797/i=2/a=L27?key=crossref.b7074258dcd51db992beb39a612431ce.

[134] A. V. Kravtsov, A. A. Vikhlinin, and A. V. Meshcheryakov. Stellar MassâĂŤHalo Mass Relation and Star Formation Efficiency in High-Mass Halos. *Astronomy Letters*, 44(1):8–34, 1 2018. ISSN 1063-7737. doi: 10.1134/S1063773717120015. URL http://www.sdss.org/dr8/http://arxiv.org/abs/1401.7329%0Ahttp://dx.doi.org/10.1134/S1063773717120015http://link.springer.com/10.1134/S1063773717120015.

[135] Charlie Conroy, James E. Gunn, Charlie Conroy, and James E. Gunn. FSPS: Flexible Stellar Population Synthesis. *ascl*, page ascl:1010.043, 2010.

[136] Charlie Conroy, Martin White, and James E. Gunn. The propagation of uncertainties in stellar population synthesis modeling. II. the challenge of comparing galaxy

evolution models to observations. *Astrophysical Journal*, 708(1):58–70, 2010. ISSN 15384357. doi: 10.1088/0004-637X/708/1/58.

[137] A. Cattaneo, A. Dekel, J. Devriendt, B. Guiderdoni, and J. Blaizot. Modelling the galaxy bimodality: Shutdown above a critical halo mass. *Monthly Notices of the Royal Astronomical Society*, 370(4):1651–1665, 2006. ISSN 00358711. doi: 10.1111/j.1365-2966.2006.10608.x.

[138] Paramita Barai, Simona Gallerani, Andrea Pallottini, Andrea Ferrara, Alessandro Marconi, Claudia Cicone, Roberto Maiolino, and Stefano Carniani. Quasar outflows at z âĽě 6: The impact on the host galaxies. *Monthly Notices of the Royal Astronomical Society*, 473(3):4003–4020, 2018. ISSN 13652966. doi: 10.1093/mnras/stx2563. URL http://arxiv.org/abs/1707.03014.

[139] Peter S Behroozi, Risa H Wechsler, and Charlie Conroy. ON THE LACK OF EVOLUTION IN GALAXY STAR FORMATION EFFICIENCY. *The Astrophysical Journal*, 762(2):L31, 1 2013. ISSN 2041-8205. doi: 10.1088/2041-8205/762/2/L31. URL http://stacks.iop.org/2041-8205/762/i=2/a=L31?key=crossref.90c62f2b48da6b365f767ee2eba8e0d7.

[140] Tom McCavana, Miroslav Micic, Geraint F Lewis, Manodeep Sinha, Sanjib Sharma, Kelly Holley-Bockelmann, and Joss Bland-Hawthorn. The lives of high-redshift mergers. *Monthly Notices of the Royal Astronomical Society*, 424(1):361–371, 7 2012. ISSN 00358711. doi: 10.1111/j.1365-2966.2012.21202.x. URL http://mnras.oxfordjournals.org/cgi/doi/10.1111/j.1365-2966.2012.21202.x.

[141] S Khochfar and A Burkert. Orbital parameters of merging dark matter halos. *Astronomy & Astrophysics*, 445(2):403–412, 2006. ISSN 0004-6361. doi: 10.1051/0004-6361:20053241. URL https://www.aanda.org/articles/aa/pdf/2006/02/aa3241-05.pdfhttp://www.aanda.org/10.1051/0004-6361:20053241.

[142] C Y Jiang, Y P Jing, A Faltenbacher, W P Lin, and Cheng Li. A Fitting Formula for the Merger Timescale of Galaxies in Hierarchical Clustering. *The Astrophysical Journal*, 675(2):1095–1105, 2008. ISSN 0004-637X. doi: 10.1086/526412. URL http://iopscience.iop.org/article/10.1086/526412/pdfhttp://stacks.iop.org/0004-637X/675/i=2/a=1095.

[143] Nicholas Lee, D B Sanders, Caitlin M Casey, Sune Toft, N Z Scoville, Chao-Ling Hung, Emeric Le Floc'h, Olivier Ilbert, H Jabran Zahid, HervÃĪ Aussel, Peter Capak, Jeyhan S Kartaltepe, Lisa J Kewley, Yanxia Li, Kevin Schawinski, Kartik Sheth, and Quanbao Xiao. A TURNOVER IN THE GALAXY MAIN SEQUENCE OF STAR FORMATION AT M * âĹij 10 10 M âŸĽ FOR REDSHIFTS z $<$ 1.3.

The Astrophysical Journal, 801(2):80, 3 2015. ISSN 1538-4357. doi: 10.1088/0004-637X/801/2/80. URL http://stacks.iop.org/0004-637X/801/i=2/a=80?key=crossref.3f3dee105084fa0b38ba3f3aaac73904.

[144] Samir Salim, R Michael Rich, Stephane Charlot, Jarle Brinchmann, Benjamin D Johnson, David Schiminovich, Mark Seibert, Ryan Mallery, Timothy M Heckman, Karl Forster, Peter G Friedman, D Christopher Martin, Patrick Morrissey, Susan G Neff, Todd Small, Ted K Wyder, Luciana Bianchi, Jose Donas, Young-Wook Lee, Barry F Madore, Bruno Milliard, Alex S Szalay, Barry Y Welsh, and Sukyoung K Yi. UV Star Formation Rates in the Local Universe. *The Astrophysical Journal Supplement Series*, 173(2):267–292, 2007. ISSN 0067-0049. doi: 10.1086/519218. URL http://iopscience.iop.org/article/10.1086/519218/pdfhttp://stacks.iop.org/0067-0049/173/i=2/a=267.

[145] Jeremy L Tinker, Brant E Robertson, Andrey V Kravtsov, Anatoly Klypin, Michael S Warren, Gustavo Yepes, and Stefan Gottlöber. The large-scale bias of dark matter halos: Numerical calibration and model tests. *Astrophysical Journal*, 724(2):878–886, 2010. ISSN 15384357. doi: 10.1088/0004-637X/724/2/878. URL http://iopscience.iop.org/article/10.1088/0004-637X/724/2/878/pdf.

[146] Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, Lars Hernquist, RÃijdiger Pakmor, Shy Genel, Paul Torrey, Mark Vogelsberger, Guinevere Kauffmann, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: The galaxy colour bimodality. *Monthly Notices of the Royal Astronomical Society*, 475(1):624–647, 2018. ISSN 13652966. doi: 10.1093/mnras/stx3040. URL http://www.tng-project.org.

[147] Volker Springel, RÃijdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, Shy Genel, Paul Torrey, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: Matter and galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 475(1):676–698, 2018. ISSN 13652966. doi: 10.1093/mnras/stx3304. URL http://www.tng-project.org.

[148] Eyal Neistein, Cheng Li, Sadegh Khochfar, Simone M Weinmann, Francesco Shankar, and Michael Boylan-kolchin. A tale of two populations : the stellar mass of central and satellite galaxies arXiv : 1103 . 3272v2 [astro-ph . CO] 30 May 2011. *Mon. Not. R. Astron. Soc*, 15(February):1–15, 2013. URL https://arxiv.org/pdf/1103.3272.pdf.

[149] Bruno M B Henriques, Simon D M White, Peter A Thomas, Raul Angulo, Qi Guo, Gerard Lemson, Volker Springel, and Roderik Overzier. Galaxy

formation in the Planck cosmology âĂŞ I. Matching the observed evolution of star formation rates, colours and stellar masses. *Monthly Notices of the Royal Astronomical Society*, 451(3):2663–2680, 8 2015. ISSN 1365-2966. doi: 10.1093/mnras/stv705. URL http://academic.oup.com/mnras/article/451/3/2663/1199296/Galaxy-formation-in-the-Planck-cosmology-Ihttps://academic.oup.com/mnras/article-abstract/451/3/2663/1199296.

[150] Gian Luigi Granato, Gianfranco De Zotti, Laura Silva, Alessandro Bressan, and Luigi Danese. A Physical Model for the Coevolution of QSOs and Their Spheroidal Hosts. *The Astrophysical Journal*, 600(2):580–594, 1 2004. ISSN 0004-637X. doi: 10.1086/379875. URL https://iopscience.iop.org/article/10.1086/379875/pdfhttp://arxiv.org/abs/astro-ph/0307202%0Ahttp://dx.doi.org/10.1086/379875http://stacks.iop.org/0004-637X/600/i=2/a=580http://iopscience.iop.org/article/10.1086/379875/pdf.

[151] A Dekel, Y Birnboim, G Engel, J Freundlich, T Goerdt, M Mumcuoglu, E Neistein, C Pichon, R Teyssier, and E Zinger. Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature*, 457 (7228):451–454, 1 2009. ISSN 0028-0836. doi: 10.1038/nature07648. URL https://arxiv.org/pdf/0808.0553.pdfhttps://www.nature.com/articles/nature07648.pdfhttp://www.nature.com/doifinder/10.1038/nature07648.

[152] Simon J Lilly, C Marcella Carollo, Antonio Pipino, Alvio Renzini, and Yingjie Peng. GAS REGULATION OF GALAXIES: THE EVOLUTION OF THE COSMIC SPECIFIC STAR FORMATION RATE, THE METALLICITY-MASS-STAR-FORMATION RATE RELATION, AND THE STELLAR CONTENT OF HALOS. *The Astrophysical Journal*, 772(2):119, 7 2013. ISSN 0004-637X. doi: 10.1088/0004-637X/772/2/119. URL https://iopscience.iop.org/article/10.1088/0004-637X/772/2/119/pdfhttp://stacks.iop.org/0004-637X/772/i=2/a=119?key=crossref.c09c66e9cf65d9f2836abac46801d4c7.

[153] Kevin Schawinski, C. Megan Urry, Brooke D Simmons, Lucy Fortson, Sug-ata Kaviraj, William C Keel, Chris J Lintott, Karen L Masters, Robert C Nichol, Marc Sarzi, Ramin Skibba, Ezequiel Treister, Kyle W Willett, O. Ivy Wong, and Sukyoung K Yi. The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxiesâŸĚ. *Monthly Notices of the Royal Astronomical Society*, 440 (1):889–907, 5 2014. ISSN 1365-2966. doi: 10.1093/mnras/stu327. URL https://arxiv.org/pdf/1402.4814.pdfhttp://academic.oup.com/mnras/article/440/1/889/1749989/The-green-valley-is-a-red-herring-Galaxy-Zoo.

[154] Pieter G. van Dokkum, Katherine E Whitaker, Gabriel Brammer, Marijn Franx, Mariska Kriek, Ivo Labbé, Danilo Marchesini, Ryan Quadri, Rachel Bezanson, Garth D Illingworth, Adam Muzzin, Gregory Rudnick, Tomer Tal, and David Wake. THE GROWTH OF MASSIVE GALAXIES SINCE z = 2. *The Astrophysical Journal*, 709(2):1018–1041, 2 2010. ISSN 0004-637X. doi: 10.1088/0004-637X/709/2/1018. URL http://www.cfht.hawaii.edu/Science/CFHLS/http://stacks.iop.org/0004-637X/709/i=2/a=1018?key=crossref.65b4539189e85b193935eab46dbbcec5.

[155] Francesco Shankar, Federico Marulli, Mariangela Bernardi, Simona Mei, Alan Meert, and Vinu Vikram. Size evolution of spheroids in a hierarchical Universe. *Mon. Not. R. Astron. Soc*, 128:109–128, 2013. doi: 10.1093/mnras/sts001. URL https://arxiv.org/pdf/1105.6043.pdf.

[156] Stewart Buchan and Francesco Shankar. Setting firmer constraints on the evolution of the most massive, central galaxies from their local abundances and ages. *Monthly Notices of the Royal Astronomical Society*, 462(2):2001–2010, 10 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw1771. URL http://arxiv.org/abs/1607.05732http://dx.doi.org/10.1093/mnras/stw1771https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw1771.

[157] DaniÃÍl N. Groenewald, Rosalind E. Skelton, David G. Gilbank, and S. Ilani Loubser. The close pair fraction of BCGs since z = 0.5: major mergers dominate recent BCG stellar mass growth. *Monthly Notices of the Royal Astronomical Society*, 467(4):4101–4117, 6 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx340. URL http://arxiv.org/abs/1701.09012https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stx340.

[158] J. Matharu, A. Muzzin, G. B. Brammer, R. F.J. Van Der Burg, M. W. Auger, P. C. Hewett, A. Van Der Wel, P. Van Dokkum, M. Balogh, J. C.C. Chan, R. Demarco, D. Marchesini, E. J. Nelson, A. Noble, G. Wilson, and H. K.C. Yee. HST/WFC3 grism observations of z âĹij 1 clusters: The cluster versus field stellar mass-size relation and evidence for size growth of quiescent galaxies from minor mergers. *Monthly Notices of the Royal Astronomical Society*, 484(1):595–617, 3 2019. ISSN 13652966. doi: 10.1093/mnras/sty3465. URL https://arxiv.org/pdf/1811.06548v1.pdfhttps://academic.oup.com/mnras/article/484/1/595/5262407.

[159] Mariangela Bernardi, Nathan Roche, Francesco Shankar, and Ravi K. Sheth. Evidence of major dry mergers at M* $>$ 2 ÃŮ 1011MâŁŹ from curvature in early-type galaxy scaling relations? *Monthly Notices of the Royal Astronomical Society: Letters*, 412(1):L6–L10, 3 2011. ISSN 17453933. doi: 10.1111/j.1745-3933.2010.00982.

x. URL https://academic.oup.com/mnrasl/article-lookup/doi/10.1111/j.1745-3933.2010.00982.x.

[160] Mark Vogelsberger, Shy Genel, Volker Springel, Paul Torrey, Debora Sijacki, Dandan Xu, Greg Snyder, Dylan Nelson, and Lars Hernquist. Introducing the illustris project: Simulating the coevolution of dark and visible matter in the universe. *Monthly Notices of the Royal Astronomical Society*, 444(2):1518–1547, 2014. ISSN 13652966. doi: 10.1093/mnras/stu1536. URL https://arxiv.org/pdf/1405.2921.pdf.

[161] Francesco Shankar. Black Hole Demography: From scaling relations to models. *Classical & Quantum Gravity*, 30(24):20, 2013. ISSN 02649381. doi: 10.1088/0264-9381/30/24/244001. URL http://arxiv.org/abs/1307.3289.

[162] A Lapi, J González-Nuevo, L Fan, A Bressan, G De Zotti, L Danese, M Negrello, L Dunne, S Eales, S Maddox, R Auld, M Baes, D G Bonfield, S Buttiglione, A Cava, D L Clements, A Cooray, A Dariush, S Dye, J Fritz, D Herranz, R Hopwood, E Ibar, R Ivison, M J Jarvis, S Kaviraj, M. López-Caniego, M Massardi, M. J. Michałowski, E Pascale, M Pohlen, E Rigby, G Rodighiero, S Serjeant, D. J.B. Smith, P Temi, J Wardlow, and P Van Der Werf. Herschel-atlas galaxy counts and high-redshift luminosity functions: The formation of massive early-type galaxies. *Astrophysical Journal*, 742(1):24, 2011. ISSN 15384357. doi: 10.1088/0004-637X/742/1/24. URL https://iopscience.iop.org/article/10.1088/0004-637X/742/1/24/pdf.

[163] M Bernardi, 1âĂź A Meert, V Vikram, M Huertas-Company, S Mei, F Shankar, and R K Sheth. Systematic effects on the sizeâĂŞluminosity relations of early-and late-type galaxies: dependence on model fitting and morphology. *MNRAS*, 443:874–897, 2014. doi: 10.1093/mnras/stu1106. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/443/1/10.1093/mnras/stu1106/2/stu1106.pdf?Expires=1498823843&Signature=SRtTC7TTJFEEJJmJJWEN~TmYtclrg9lw5mfu3jQ0DP5aZkbBzNtRr25Conh20h744tKGTzAJDEByNeBDtMamNJI8

[164] Michele Cappellari, Richard M. McDermid, Katherine Alatalo, Leo Blitz, Maxime Bois, FrÃľdÃľric Bournaud, M Bureau, Alison F Crocker, Roger L Davies, Timothy A Davis, P. T. de Zeeuw, Pierre Alain Duc, Eric Emsellem, Sadegh Khochfar, Davor Krajnović, Harald Kuntschner, Raffaella Morganti, Thorsten Naab, Tom Oosterloo, Marc Sarzi, Nicholas Scott, Paolo Serra, Anne Marie Weijmans, and Lisa M Young. The ATLAS3D project-XX. Mass-size and mass-σ distributions of early-type galaxies: Bulge fraction drives kinematics, mass-to-light ratio, molecular gas fraction and stellar initial mass function. *Monthly Notices of the Royal Astronomical Society*, 432(3):1862–1893, 2013. ISSN 00358711. doi: 10.1093/mnras/stt644. URL http://purl.org/cappellari/idl.

[165] Michele Cappellari. EFFECT OF ENVIRONMENT ON GALAXIES' MASS-SIZE DISTRIBUTION: UNVEILING THE TRANSITION FROM OUTSIDE-IN TO INSIDE-OUT EVOLUTION. *The Astrophysical Journal*, 778 (1):L2, 10 2013. ISSN 2041-8205. doi: 10.1088/2041-8205/778/1/L2. URL https://arxiv.org/pdf/1309.1136.pdfhttp://stacks.iop.org/2041-8205/778/i=1/a=L2?key=crossref.29f068d22b9a8c8b93499056ac8f631d.

[166] Adam R. Tomczak, Ryan F. Quadri, Kim Vy H. Tran, Ivo Labbé, Caroline M.S. Straatman, Casey Papovich, Karl Glazebrook, Rebecca Allen, Gabriel B. Brammer, Glenn G. Kacprzak, Lalitwadee Kawinwanichakij, Daniel D. Kelson, Patrick J. McCarthy, Nicola Mehrtens, Andrew J. Monson, S. Eric Persson, Lee R. Spitler, Vithal Tilvi, and Pieter Van Dokkum. Galaxy stellar mass functions from zfourge/-candels: An excess of low-mass galaxies since z = 2 and the rapid buildup of quiescent galaxies. *Astrophysical Journal*, 783(2), 2014. ISSN 15384357. doi: 10.1088/0004-637X/783/2/85. URL http://iopscience.iop.org/article/10.1088/0004-637X/783/2/85/pdf.

[167] Philip J. Grylls, F. Shankar, J Leja, N. Menci, B. Moster, P. Behroozi, and L. Zanisi. Predicting fully self-consistent satellite richness, galaxy growth and starformation rates from the STastical sEmi-Empirical modeL steel. *Monthly Notices of the Royal Astronomical Society*, 000:1–19, 10 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz2956. URL http://arxiv.org/abs/1910.08417http://dx.doi.org/10.1093/mnras/stz2956https://academic.oup.com/mnras/advance-article/doi/10.1093/mnras/stz2956/5602609.

[168] F. Fiore, C. Feruglio, F. Shankar, M. Bischetti, A. Bongiorno, M. Brusa, S. Carniani, C. Cicone, F. Duras, A. Lamastra, V. Mainieri, A. Marconi, N. Menci, R. Maiolino, E. Piconcelli, G. Vietri, and L. Zappacosta. AGN wind scaling relations and the co-evolution of black holes and galaxies. *Astronomy and Astrophysics*, 601, 2017. ISSN 14320746. doi: 10.1051/0004-6361/201629478.

[169] J. Negroponte and S. D. M. White. Simulations of mergers between disc-halo galaxies. *Monthly Notices of the Royal Astronomical Society*, 205(4):1009–1029, 12 1983. ISSN 0035-8711. doi: 10.1093/mnras/205.4.1009. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/205.4.1009.

[170] Christopher J. Conselice, Matthew A. Bershady, Mark Dickinson, and Casey Papovich. A Direct Measurement of Major Galaxy Mergers at z 3 . *The Astronomical Journal*, 126(3):1183–1207, 2003. ISSN 0004-6256. doi: 10.1086/377318. URL https://arxiv.org/pdf/astro-ph/0306106.pdf.

[171] Christopher J Conselice, Sheena Rajgor, and Robert Myers. The structures of distant galaxies - I. Galaxy structures and the merger rate to z âĹij 3 in the Hubble Ultra-Deep Field. *Monthly Notices of the Royal Astronomical Society*, 386 (2):909–927, 2008. ISSN 00358711. doi: 10.1111/j.1365-2966.2008.13069.x. URL https://academic.oup.com/mnras/article-abstract/386/2/909/1058396.

[172] Kenneth Duncan, Christopher J Conselice, Carl Mundy, Eric Bell, Jennifer Donley, Audrey Galametz, Yicheng Guo, Norman A Grogin, Nimish Hathi, Jeyhan Kartaltepe, Dale Kocevski, Anton M Koekemoer, Pablo G Pérez-González, Kameswara B Mantha, Gregory F Snyder, and Mauro Stefanon. Observational Constraints on the Merger History of Galaxies since z âĽĹ 6: Probabilistic Galaxy Pair Counts in the CANDELS Fields. *The Astrophysical Journal*, 876(2):110, 2019. ISSN 1538-4357. doi: 10.3847/1538-4357/ab148a.

[173] Pieter G. van Dokkum, Joel Leja, Erica June Nelson, Shannon Patel, Rosalind E. Skelton, Ivelina Momcheva, Gabriel Brammer, Katherine E. Whitaker, Britt Lundgren, Mattia Fumagalli, Charlie Conroy, Natascha Förster Schreiber, Marijn Franx, Mariska Kriek, Ivo Labbé, Danilo Marchesini, Hans-Walter Rix, Arjen van der Wel, and Stijn Wuyts. The Assembly of Milky-Way-like Galaxies Since z ˜ 2.5. *The Astrophysical Journal*, 771(2):L35, 7 2013. ISSN 0004-637X. doi: 10.1088/2041-8205/771/2/L35.

[174] M. Huertas-Company, M. Bernardi, P. G. Pérez-González, M. L N Ashby, G. Barro, C. Conselice, E. Daddi, A. Dekel, P. Dimauro, S. M. Faber, N. A. Grogin, J. S. Kartaltepe, D. D. Kocevski, A. M. Koekemoer, D. C. Koo, S. Mei, and F. Shankar. Mass assembly and morphological transformations since $z\sim3$ from CANDELS. *Monthly Notices of the Royal Astronomical Society*, 6 2016. ISSN 13652966. doi: 10.1093/mnras/stw1866. URL http://arxiv.org/abs/1606.04952http://dx.doi.org/10.1093/mnras/stw1866.

[175] Joel Leja, Pieter van Dokkum, and Marijn Franx. TRACING GALAXIES THROUGH COSMIC TIME WITH NUMBER DENSITY SELECTION. *The Astrophysical Journal*, 766(1):33, 3 2013. ISSN 0004-637X. doi: 10.1088/0004-637X/766/1/33. URL https://iopscience.iop.org/article/10.1088/0004-637X/766/1/33/pdfhttp://stacks.iop.org/0004-637X/766/i=1/a=33?key=crossref.d050fc2302dc1459668a88202bfa5cb0.

[176] Carl J Mundy, Christopher J Conselice, and Jamie R Ownsworth. Tracing galaxy populations through cosmic time: A critical test of methods for connecting the same galaxies between different redshifts at z $<$ 3. *Monthly Notices of the Royal Astronomical Society*, 450(4):3696–3707, 2015. ISSN 13652966. doi: 10.1093/mnras/stv860. URL http://www.mpa-garching.mpg.de/galform/virgo/millennium/.

[177] F Bournaud, J. Jog, and F Combes. Multiple minor mergers: Formation of elliptical galaxies and constraints for the growth of spiral disks. *Astronomy and Astrophysics*, 476(3):1179–1190, 2007. ISSN 00046361. doi: 10.1051/0004-6361:20078010.

[178] Philip F Hopkins, Rachel S Somerville, Thomas J Cox, Lars Hernquist, Shardha Jogee, Dusan Kereš, Chung Pei Ma, Brant Robertson, and Kyle Stewart. The effects of gas on morphological transformation in mergers: Implications for bulge and disc demographics. *Monthly Notices of the Royal Astronomical Society*, 397(2):802–814, 8 2009. ISSN 00358711. doi: 10.1111/j.1365-2966.2009.14983.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/397/2/10.1111/j.1365-2966.2009.14983.x/2/mnras0397-0802.pdf?Expires=1496511765&Signature=Sljh7Pc18CSy7PMOC1vN5DySwv7tskTaQ4PblhpQFD51VkRAp5PrA68Gl5OvdqEcK8MtJhIsgd8hlKcvmd

[179] Francesco Shankar, Federico Marulli, Mariangela Bernardi, Simona Mei, Alan Meert, and Vinu Vikram. Size Evolution of Spheroids in a Hierarchical Universe. *Monthly Notices of the Royal Astronomical Society*, 428(1):109–128, 5 2011. ISSN 00358711. doi: 10.1093/mnras/sts001. URL http://arxiv.org/abs/1105.6043http://dx.doi.org/10.1093/mnras/sts001.

[180] Fabio Fontanot, Andrea V. Macciò, Michaela Hirschmann, Gabriella De Lucia, Rahul Kannan, Rachel S Somerville, and Dave Wilman. On the dependence of galaxy morphologies on galaxy mergers. *Monthly Notices of the Royal Astronomical Society*, 451(3):2968–2977, 2015. ISSN 13652966. doi: 10.1093/mnras/stv1119. URL https://arxiv.org/pdf/1507.04748.pdf.

[181] C M Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69(12):3101–3156, 12 2006. ISSN 0034-4885. doi: 10.1088/0034-4885/69/12/R02. URL http://iopscience.iop.org/article/10.1086/379875/pdfhttp://stacks.iop.org/0004-637X/600/i=2/a=580http://stacks.iop.org/0034-4885/69/i=12/a=R02?key=crossref.09ba099291f3712fe6016f7b1bfda1ef.

[182] Rowena K Malbon, C. M. Baugh, C S Frenk, and C G Lacey. Black hole growth in hierarchical galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 382(4):1394–1414, 2007. ISSN 00358711. doi: 10.1111/j.1365-2966.2007.12317.x. URL http://articles.adsabs.harvard.edu/pdf/2007MNRAS.382.1394M.

[183] R G Bower, I Vernon, M Goldstein, A J Benson, C G Lacey, C M Baugh, S Cole, and C S Frenk. The parameter space of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 407(4):2017–2045, 2010. ISSN 00358711. doi: 10.1111/j.1365-2966.2010.16991.x. URL https://arxiv.org/pdf/1004.0711.pdf.

[184] Philip F Hopkins, Lars Hernquist, Thomas J Cox, Tiziana Di Matteo, Brant Robertson, and Volker Springel. A Unified, MergerâĂŘdriven Model of the Origin of Starbursts, Quasars, the Cosmic XâĂŘRay Background, Supermassive Black Holes, and Galaxy Spheroids. *The Astrophysical Journal Supplement Series*, 163(1):1–49, 2006. ISSN 0067-0049. doi: 10.1086/499298. URL http://iopscience.iop.org/article/10.1086/499298/pdf.

[185] Philip F Hopkins, Darren Croton, Kevin Bundy, Sadegh Khochfar, Frank van den Bosch, Rachel S Somerville, Andrew Wetzel, Dusan Keres, Lars Hernquist, Kyle Stewart, Joshua D Younger, Shy Genel, and Chung-Pei Ma. MERGERS IN ΛCDM: UNCERTAINTIES IN THEORETICAL PREDICTIONS AND INTERPRETATIONS OF THE MERGER RATE. *The Astrophysical Journal*, 724(2):915–945, 12 2010. ISSN 0004-637X. doi: 10.1088/0004-637X/724/2/915. URL http://iopscience.iop.org/article/10.1088/0004-637X/724/2/915/pdfhttp://stacks.iop.org/0004-637X/724/i=2/a=915?key=crossref.0397dcee41fc227da84d86f424b1abdc.

[186] Joseph A. O'Leary, Benjamin P. Moster, Thorsten Naab, and Rachel S. Somerville. EMERGE: Empirical predictions of galaxy merger rates since $z\sim6$. 1 2020. URL http://arxiv.org/abs/2001.02687.

[187] Kyle R. Stewart, James S. Bullock, Risa H. Wechsler, Ariyeh H. Maller, and Andrew R. Zentner. Merger Histories of Galaxy Halos and Implications for Disk Survival. *The Astrophysical Journal*, 683(2):597–610, 8 2008. ISSN 0004-637X. doi: 10.1086/588579. URL http://stacks.iop.org/0004-637X/683/i=2/a=597.

[188] Kyle R Stewart, James S Bullock, Elizabeth J Barton, and Risa H Wechsler. GALAXY MERGERS AND DARK MATTER HALO MERGERS IN ΛCDM: MASS, REDSHIFT, AND MASS-RATIO DEPENDENCE. *The Astrophysical Journal*, 702(2):1005–1015, 9 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/702/2/1005. URL http://stacks.iop.org/0004-637X/702/i=2/a=1005?key=crossref.64fb4e87004fbda620397d058b0acc54.

[189] P CHAMARAUX, C BALKOWSKI, and P FONTANELLI. The H I content of lenticular and early-type galaxies: a comparison between field and Virgo cluster samples. *Astronomy and astrophysics (Berlin. Print)*, 165(1):15–30, 1986. ISSN 0004-6361.

[190] Alan Dressler. Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *The Astrophysical Journal*, 236:351, 3 1980. ISSN 0004-637X. doi: 10.1086/157753. URL http://adsabs.harvard.edu/doi/10.1086/157753.

[191] M Postman, M Franx, N J G Cross, B Holden, H C Ford, G D Illingworth, T Goto, R Demarco, P Rosati, J P Blakeslee, K.âĂŘV. Tran, N. Benitez, M Clampin, G F Hartig, N Homeier, D R Ardila, F Bartko, R J Bouwens, L D Bradley, T J Broadhurst, R A Brown, C J Burrows, E S Cheng, P D Feldman, D A Golimowski, C Gronwall, L Infante, R A Kimble, J E Krist, M P Lesser, A R Martel, S Mei, F Menanteau, G R Meurer, G K Miley, V Motta, M Sirianni, W B Sparks, H D Tran, Z I Tsvetanov, R L White, and W. Zheng. The MorphologyâĂŘDensity Relation in z âĹij 1 Clusters. *The Astrophysical Journal*, 623(2):721–741, 4 2005. ISSN 0004-637X. doi: 10.1086/428881. URL http://stacks.iop.org/0004-637X/623/i=2/a=721.

[192] M. Huertas-Company, P. G. Pérez-González, S. Mei, F. Shankar, M. Bernardi, E. Daddi, G. Barro, G. Cabrera-Vives, A. Cattaneo, P. Dimauro, and R. Gravet. The morphologies of massive galaxies from z âĹij 3 - Witnessing the two channels of bulge growth. *Astrophysical Journal*, 809(1), 2015. ISSN 15384357. doi: 10.1088/0004-637X/809/1/95. URL https://arxiv.org/pdf/1506.03084.pdf.

[193] Philip F Hopkins, Lars Hernquist, Thomas J Cox, Dusan Keres, and Stijn Wuyts. DISSIPATION AND EXTRA LIGHT IN GALACTIC NUCLEI. IV. EVOLUTION IN THE SCALING RELATIONS OF SPHEROIDS. *The Astrophysical Journal*, 691(2):1424–1458, 2 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/691/2/1424. URL http://iopscience.iop.org/article/10.1088/0004-637X/691/2/1424/pdfhttp://stacks.iop.org/0004-637X/691/i=2/a=1424?key=crossref.c8f49366dd1ee35f5148c5bc1119849b.

[194] Frederic Bournaud, Avishai Dekel, Romain Teyssier, Marcello Cacciato, Emanuele Daddi, Stephanie Juneau, and Francesco Shankar. Black Hole growth and AGN obscuration by instability-driven inflows in high-redshift disk galaxies fed by cold streams. *Astrophysical Journal Letters*, 741(2), 7 2011. ISSN 20418205. doi: 10.1088/2041-8205/741/2/L33. URL https://arxiv.org/pdf/1107.1483.pdfhttp://arxiv.org/abs/1107.1483http://dx.doi.org/10.1088/2041-8205/741/2/L33.

[195] Eija Laurikainen, Heikki Salo, and Ronald Buta. Multicomponent decompositions for a sample of S0 galaxies. *Monthly Notices of the Royal Astronomical Society*, 362(4):1319–1347, 2005. ISSN 00358711. doi: 10.1111/j.1365-2966.2005.09404.x.

[196] Sidney Van Den Bergh. Luminosities of barred and unbarred S0 galaxies. *Astrophysical Journal*, 754(1), 5 2012. ISSN 15384357. doi: 10.1088/0004-637X/754/1/68. URL http://arxiv.org/abs/1205.6183http://dx.doi.org/10.1088/0004-637X/754/1/68.

[197] Andrew P. Hearin, Andrew R. Zentner, Andrew A. Berlind, and Jeffrey A. Newman. SHAM beyond clustering: New tests of galaxy-halo abundance matching with galaxy groups. Technical Report 1, 2013. URL http://www.multidark.org.

[198] K. M. Hambleton, B. K. Gibson, C. B. Brook, G. S. Stinson, C. J. Conselice, J. Bailin, H. Couchman, and J. Wadsley. Advanced morphological galaxy classification: A comparison of observed and simulated galaxies. *Monthly Notices of the Royal Astronomical Society*, 418(2):801–810, 12 2011. ISSN 00358711. doi: 10.1111/j.1365-2966.2011.19532.x. URL http://arxiv.org/abs/1107.6045http://dx.doi.org/10.1111/j.1365-2966.2011.19532.x.

[199] F. Song, S. Parekh, L. Hooper, Y. K. Loke, J. Ryder, A. J. Sutton, C. Hing, C. S. Kwok, C. Pang, and I. Harvey. Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment*, 14(8):1–220, 2 2010. ISSN 13665278. doi: 10.3310/hta14080.

[200] Fabio Fontanot, Pierluigi Monaco, Laura Silva, and Andrea Grazian. Reproducing the assembly of massive galaxies within the hierarchical cosmogony. *Monthly Notices of the Royal Astronomical Society*, 382(2):903–914, 12 2007. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2007.12449.x. URL https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mnras/382/2/10.1111/j.1365-2966.2007.12449.x/2/mnras0382-0903.pdf?Expires=1495893786&Signature=TORj92V68LbSwcuXCZpRs1W3wRIM7~S9qciWEJj3rTAAEi-~flhd6Z7-N83A5ALRvxhRY6fzNsopZMASe6i3BRgZ

[201] Leo Blitz and Erik Rosolowsky. The Role of Pressure in GMC Formation II: The H 2 âĂŘPressure Relation. *The Astrophysical Journal*, 650(2):933–944, 2006. ISSN 0004-637X. doi: 10.1086/505417.

[202] Mark R Krumholz, Christopher F. McKee, and Jason Tumlinson. The AtomicâĂŘtoâĂŘMolecular Transition in Galaxies. I. An Analytic Approximation for Photodissociation Fronts in Finite Clouds. *The Astrophysical Journal*, 689(2):865–882, 12 2008. ISSN 0004-637X. doi: 10.1086/592490. URL http://arxiv.org/abs/0805.2947http://dx.doi.org/10.1086/592490http://stacks.iop.org/0004-637X/689/i=2/a=865.

[203] Mark R Krumholz, Christopher F. McKee, and Jason Tumlinson. THE ATOMIC-TO-MOLECULAR TRANSITION IN GALAXIES. II: H I AND H 2 COLUMN DENSITIES. *The Astrophysical Journal*, 693(1):216–235, 3 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/693/1/216. URL http://stacks.iop.org/0004-637X/693/i=1/a=216?key=crossref.0df05158d6c91e7468d4726516534bad.

[204] Mark R Krumholz, Christopher F. McKee, and Jason Tumlinson. THE STAR FORMATION LAW IN ATOMIC AND MOLECULAR GAS. *The Astrophysical Journal*, 699(1):850–856, 7 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/699/1/850. URL http://arxiv.org/abs/0904.0009http://dx.doi.org/10.1088/0004-637X/699/1/850http://stacks.iop.org/0004-637X/699/i=1/a=850?key=crossref.f4780a03a7548ab10c1a27abf85409e6.