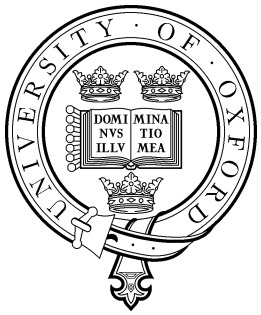
The Relationship Between Galaxies and Their Dark Matter Haloes Over Cosmic Time



Peter William Hatfield
Lincoln College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2017

Abstract

In this thesis I study and measure the spatial distribution of galaxies selected in optical and near-infrared surveys over cosmic time. By measuring the clustering of these sources, valuable insight can be gained into the role of environment in shaping galaxy evolution over the history of the Universe.

I present a series of results from a clustering analysis of the first data release of the Visible and Infrared Survey Telescope for Astronomy (VISTA) Deep Extragalactic Observations (VIDEO) survey. VIDEO is the only survey currently capable of probing the bulk of stellar mass in galaxies at redshifts corresponding to the peak of star formation on degree scales. Galaxy clustering is measured with the two-point correlation function, which is calculated using a non parametric kernel based density estimator. I use my measurements to investigate the connection between the galaxies and the host dark matter halo using a Halo Occupation Distribution (HOD) methodology, deriving bias, satellite fractions, and typical host halo masses for stellar masses between $10^{9.35}M_{\odot}$ and $10^{10.85}M_{\odot}$, at redshifts $0.5 < z < 1.7$. I show that the typical halo mass increases with stellar mass (with moderate scatter) and bias also increases with stellar mass and redshift, consistent with previous studies. I find the satellite fraction increases towards low redshifts, from $\sim 5\%$ at $z \sim 1.5$, to $\sim 20\%$ at $z \sim 0.6$, also increasing for lower mass galaxies. I combine my results to derive the stellar mass to halo mass ratio for both satellites and centrals over a range of halo masses and find the peak corresponding to the halo mass with maximum star formation efficiency to be $\sim 2 \times 10^{12}M_{\odot}$, finding no evidence for evolution.

It has long been known that environment has a large effect on star formation in galaxies. There are several known plausible mechanisms to remove the cool gas needed for star formation, such as strangulation, harassment and ram-pressure stripping. It is unclear which process is dominant, and over what range of stellar mass. In this thesis, I find evidence for suppression of the cross-correlation function between massive galaxies and less massive star-forming galaxies, giving a measure of how less likely a galaxy is to be star-forming in the vicinity of a more massive galaxy. I develop a formalism for modelling environmental quenching mechanisms within the HOD formalism. I find that at $z \sim 2$ environment is not a significant factor in determining quenching of star-forming galaxies, and that galaxies are quenched with similar probabilities in group environments as they are globally. However, by $z \sim 0.5$ galaxies are much less likely to be star forming when in a group environment than when not. This increased probability of being quenched does not appear to have significant radial dependence within the halo, supportive of the quenching being caused by the halting of fresh inflows of pristine gas, as opposed to by tidal stripping. Furthermore, by separating the massive sample into passive and star-forming, I find that this effect is

further enhanced when the central galaxy is passive, a manifestation of galactic conformity.

Hydrodynamical cosmological simulations, with advances in computing power over the last decade, have recently made great advances in reproducing the galaxy population and understanding the underlying physical processes behind galaxy evolution. There is extensive research in the literature comparing predicted stellar mass functions from hydrodynamical simulations to observed stellar mass functions in data. In this thesis I extend these results, comparing clustering of galaxies in mock catalogues from the hydrodynamical cosmological simulation Horizon-AGN to clustering measurements from the VIDEO observations. Clustering and HOD modelling in the Horizon-AGN mock catalogue qualitatively recreates clustering measurements from the VIDEO data, but reflects the known excess stellar mass to halo mass ratio for low mass haloes in Horizon-AGN. This reinforces the need for stronger regulation of star formation in low mass haloes in the simulation.

I extend my results into the high redshift regime by studying the large-scale structure of the bright high-redshift Lyman-break galaxy (LBG) population - gaining insight into the role of environment in galaxy formation physics in the early Universe. I measure the clustering of a sample of bright ($-22.7 < M_{UV} < -21.125$) LBGs at $z \sim 6$ and use a HOD model to measure their typical halo masses. I find that the clustering amplitude and corresponding HOD fits suggests that these sources are highly biased ($b \sim 10$) objects in the densest regions of the high-redshift Universe. Coupled with the observed rapid evolution of the number density of these objects, my results suggest that the shape of high luminosity end of the luminosity function is related to feedback processes or the onset of dust obscuration - as opposed to a scenario where these sources are predominantly rare instances of the much more numerous $M_{UV} \sim -19$ population of galaxies caught in a particularly vigorous period of star formation. Despite investigating several variations on the model, it was not possible to simultaneously fit both the number densities and clustering measurements. I interpret this as a signal that a refinement of the model halo bias relation at high redshifts or the incorporation of quasi-linear effects may be needed for future attempts at modelling the clustering and number counts. Finally, the difference in number density between the fields (UltraVISTA has a surface density ~ 1.8 times greater than UDS) is shown to be consistent with the cosmic variance implied by the clustering measurements.

Finally I discuss future data sets that will become available in the coming years, and future approaches to modelling large-scale structure. In summary I have shown that measuring the spatial distribution of galaxies on large-scales is a vital probe of galaxy evolution and an essential tool for understanding the connection between galaxies and their dark matter haloes over cosmic time.

Declaration

I carried out the work presented in thesis at the Department of Astrophysics, University of Oxford between October 2013 and March 2017, under the supervision of Professor Matt Jarvis. It was funded by a Science Technology Facilities Council Studentship. I hereby declare that no part of this thesis has been submitted in support of another degree, diploma or other qualification at the University of Oxford or other higher learning institute. Except where otherwise stated or where reference is made to the work of others, the work in this thesis is entirely my own.

The work in parts of Chapter 2 and all of Chapter 3 is based on the peer-reviewed, published paper Hatfield et al. 2016, MNRAS, 459, 2618, for which I am lead author, working with Sam Lindsay, Matt Jarvis, Boris Häußler, Mattia Vaccari and Aprajita Verma. The work in Chapter 4 is based on the MNRAS submitted paper Hatfield and Jarvis et al. 2016, arXiv:1606.08989, for which I am lead author, working with Matt Jarvis. Chapter 5 is based on the MNRAS submitted paper Hatfield et al., 2017 arXiv:1702.03309, for which I am the lead author, working with Rebecca Bowler, Matt Jarvis, and Catherine Hale. Chapter 6 is work carried out in collaboration with Clotilde Laigle and Matt Jarvis, due to be submitted to MNRAS shortly.

Work in this thesis is based on data products from observations made with ESO Telescopes at the La Silla or Paranal Observatories under ESO programme ID 179.A- 2006. This thesis is also based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

I would like to acknowledge support from Rector and Fellows of Lincoln College through the Graduate Research Fund. Some of this research was supported by the Oxford Centre for Astrophysical Surveys which is funded through generous support from the Hintze Family Charitable Foundation, the award of the STFC consolidated grant (ST/N000919/1), and the John Fell Oxford University Press (OUP) Research Fund. I also acknowledge funding from the Newton Fund for exchange visits between Oxford and South Africa.

This thesis made extensive use of the following codes and software: the Tool for Operations on Catalogues And Tables (TOPCAT; Taylor, 2005), the Python module Astropy (Astropy Collaboration et al., 2013), the python modules Halomod and HMF (Murray et al., 2013), and the python modules Corner and emcee (Foreman-Mackey et al., 2013). This research made use of NASAs ADS service and Cornells arXiv e-print depository. This thesis was prepared using the typesetting tool LaTeX.

Peter Hatfield (*March 2017*)

Acknowledgements

The biggest thanks and credit for my thesis goes to my supervisor Matt Jarvis who has made my DPhil an extremely fun (and hopefully occasionally productive) three and a bit years. His mentorship and encouragement has been invaluable, both scientifically and professionally, and he's given me many opportunities not least of which have been sending me around the world and involving me in fantastic projects. I have been extremely fortunate to have had such a great advisor. My only hope is that one day I might know enough about sport to at least feign a passable sporting conversation.

A big thank you to my family for encouraging me through 23 terms of tertiary education. In particular thanks to my mother for doing maths puzzles with me from a young age and transporting my stuff to and from University more than a dozen times over the last 8 years, my father for setting my sights high (even if it ended up being space, not aeroplanes), Alex (my literal El Dude Brother) for always being there for mischief and/or wise words, and to all of my grandparents for 25 years of support.

Thanks to all the great friends I've made in Oxford, in particular to Alfie Tiley, Andre Hector, Ben Pope and Simon Zieleniewski. Many happy memories of playful capers at the Fortress, Lincoln MCR, the ALP, and Purple Turtle.

Thanks to everyone else in Oxford who I've had good times with, have worked with and have helped me with my thesis, in particular Aprajita, Rebecca, Catherine, Leah M, Zahra, Holly, Boris, Sarah, Sam and all my office mates from 650 and the Tower. Special mention of course to Ashling for smoothing the gears of everyones DPhil's and life in general. Thank you also to everyone at the University of Western Cape for hosting me for 6 months, in particular Mario Santos, Matt P, Nikhita, Siyambonga, Kim, Imogen and Mattia. Thanks also to John Stott, Clotilde Laigle, Julien Devriendt, Jeremy Blaizot, Jean Coupon and Steven Murray for useful discussions.

A major thank you to Professor Becky Parker and Professor Steven Rose of course for getting me into physics in the very first place and kindling my early proto-career all those years ago!

Finally thank you to Leah Trueblood who has been my partner in crime on many adventures (academic and otherwise) over the last three years and has been the biggest support pushing through the final few months of the DPhil. Thanks for coming to the Physics Flash talks in Michaelmas 2013 for what has been by far my most successful outreach talk ever.

Peter Hatfield (*March 2017*)

Contents

1 Introduction **1**

1.1 Cosmology . . . 1
1.1.1 Big Bang Cosmology . . . 2
1.1.2 ΛCDM . . . 5
1.1.3 Redshift as time and distance measure . . . 6
1.1.3.1 Physical and Comoving Distance . . . 7
1.1.3.2 Angular and Luminosity Distance . . . 8
1.1.3.3 Look-Back Time . . . 8
1.1.4 Cosmological Probes . . . 9
1.1.4.1 Supernovae and Standard Candles . . . 9
1.1.4.2 Cosmic Microwave Background . . . 11
1.1.4.3 Baryonic Acoustic Oscillations and Standard Rulers . . . 14
1.1.4.4 Weak Lensing of Large-Scale Structure . . . 17
1.1.4.5 Joint Constraints . . . 18
1.1.5 The Nature of Dark Matter . . . 20
1.1.5.1 Evidence for Dark Matter . . . 20
1.1.5.2 Different Dark Matter Candidates . . . 21
1.2 Structure Formation . . . 24
1.2.1 Linear Perturbation Theory . . . 24
1.2.2 The Halo Model . . . 27
1.2.2.1 Spherical Collapse Model . . . 27
1.2.2.2 Press-Schechter Theory - Halo Mass Function . . . 29
1.2.2.3 Peak Background Split - Halo Bias . . . 31
1.2.2.4 N-Body Simulations . . . 33
1.2.2.5 Halo Profiles . . . 34
1.3 Galaxy Evolution . . . 35
1.3.1 Anatomy of a Galaxy Spectrum . . . 36
1.3.1.1 Ultra-Violet, Optical and Infrared . . . 37
1.3.1.2 Radio . . . 38
1.3.1.3 X-ray . . . 38
1.3.1.4 Stellar Population Synthesis Models . . . 38
1.3.2 Galaxy Formation in Halos . . . 39
1.3.3 Stellar mass function . . . 39
1.3.4 Red-Blue Bimodality . . . 41
1.3.5 Galaxies Over Cosmic Time . . . 42
1.3.6 Active Galactic Nuclei and Feedback Processes . . . 43
1.4 Wide Field Surveys . . . 43

1.4.1 Imaging . . . 44
1.4.2 Spectroscopic . . . 44

2 Measuring and Modelling Large-Scale Structure 47
2.1 The Correlation function . . . 47
2.1.1 Spatial, Projected and Angular Correlation Functions . . . 48
2.1.1.1 The Redshift-Space Distorted Spatial Correlation Function 48
2.1.1.2 Integrating to the Projected Correlation Function . . . 50
2.1.1.3 Projecting to the Angular Correlation Function . . . 51
2.1.1.4 Transforming to the Power Spectrum . . . 52
2.1.2 Calculating the Angular Correlation Function . . . 53
2.1.2.1 Error Bars . . . 55
2.1.2.2 The Integral Constraint . . . 55
2.1.2.3 Continuous Estimation of the Correlation Function . . . 57
2.1.2.4 Accounting for Redshift Probability Distribution Functions 59
2.2 Abundance Matching and Power Law Modelling of the Correlation Function 61
2.2.1 Abundance Matching . . . 61
2.2.2 Power Law Modelling of the Correlation Function . . . 61
2.3 Halo Occupation Distribution Modelling . . . 62
2.3.1 Physics Probed . . . 63
2.3.2 Derived Parameters . . . 65
2.3.3 The 2-halo Term . . . 65
2.3.4 The 1-halo Term . . . 66
2.3.4.1 The Central-Satellite Term . . . 66
2.3.4.2 The Satellite-Satellite Term . . . 67
2.3.5 Common Parametrisations . . . 67
2.3.6 Incorporating Stellar Mass Ranges in the Model . . . 70

3 The Galaxy-Halo Connection in the VIDEO Survey 72
3.1 Introduction . . . 72
3.2 Observations . . . 76
3.2.1 VIDEO and CFHTLS . . . 76
3.2.2 LePHARE and SExtractor . . . 77
3.2.3 Final Sample . . . 78
3.3 The Two-Point Correlation Function . . . 79
3.4 Halo Occupation Distribution Modelling . . . 80
3.4.1 The Model . . . 80
3.4.2 Projecting from $3D$ and Choice of $N(z)$ . . . 81
3.4.3 MCMC Fitting Process . . . 82
3.5 Results . . . 83
3.5.1 Redshift and Stellar Mass Selection . . . 83
3.5.2 $M_{\rm min}$, M_1 and the Mass Gap . . . 90
3.5.3 α and σ . . . 91
3.5.4 Insight into Substructure . . . 93
3.5.5 Derived Parameters . . . 93
3.5.6 Stellar Mass to Halo Mass Ratio . . . 94
3.6 Conclusions . . . 99

4 Environmental Quenching and Galactic Conformity in the Galaxy Cross-Correlation Signal 101
4.1 Introduction . . . 101
4.2 Observations . . . 104
4.3 The Angular Cross Correlation Function . . . 107
4.3.1 Background . . . 107
4.3.2 Application to Environmental Quenching and Conformity . . . 110
4.4 Results . . . 112
4.4.1 The High- to Low-mass Cross-Correlation Function . . . 112
4.4.2 Selection by SFR in the Massive Sample . . . 115
4.4.3 Modelling Quenching Effects Directly . . . 116
4.4.3.1 Radial Dependence of Quenching Within a Halo . . . 117
4.4.3.2 The Halo Model . . . 125
4.4.3.3 Fitting Quenching Models . . . 126
4.4.3.4 Results of the quenching model . . . 128
4.5 Discussion . . . 133
4.5.1 Summary of Results . . . 133
4.5.2 Quenching over cosmic time . . . 135
4.5.3 Future Development of the Approach . . . 140
4.6 Conclusions . . . 141

5 The environment and host haloes of the brightest $z \sim 6$ LBGs 143
5.1 Introduction . . . 143
5.1.1 Lyman-break Galaxies . . . 143
5.1.2 Clustering . . . 145
5.1.3 Probes of Reionization . . . 146
5.1.4 Objectives . . . 147
5.2 Data and Sample Selection . . . 148
5.2.1 UltraVISTA/COSMOS . . . 148
5.2.2 UDS . . . 148
5.2.3 Candidate Selection . . . 149
5.3 Correlation Functions and HOD Modelling . . . 149
5.3.1 Halo Occupation Distribution modelling . . . 150
5.3.2 Duty Cycle . . . 151
5.3.3 MCMC Fitting . . . 152
5.4 Results . . . 153
5.4.1 Clustering Measurements . . . 153
5.4.2 Modelling Results . . . 155
5.5 Discussion . . . 160
5.5.1 The link between low- and high-luminosity galaxies and their haloes at $z \sim 6$ . . . 160
5.5.2 Apparent lack of a 1-Halo term . . . 162
5.5.3 Mismatch between the number counts and bias measurements . . . 164
5.5.3.1 Contaminants . . . 164
5.5.3.2 More complex galaxy-halo relations . . . 165
5.5.3.3 Uncertainty in knowledge of the high-redshift dark matter distribution . . . 166

5.5.3.4 Modification of our understanding of high redshift structure formation . . . 167
5.5.4 Estimating Cosmic Variance . . . 167
5.5.5 The onset of quenching . . . 171
5.6 Conclusions . . . 173

6 Simulating the Universe **175**
6.1 Introduction . . . 175
6.2 Simulation Data . . . 178
6.2.1 Horizon-AGN . . . 179
6.2.2 Construction of Mock Skies . . . 179
6.2.3 Measuring Clustering . . . 180
6.3 Results . . . 181
6.3.1 ACF Measurements in Horizon-AGN . . . 181
6.3.2 Comparing VIDEO and Horizon-AGN . . . 182
6.3.3 HOD Fits . . . 183
6.4 Discussion . . . 185
6.4.1 Observational Biases . . . 185
6.4.2 Over Production of Lower Mass Galaxies in Horizon-AGN . . . 187
6.4.3 Cosmic Variance . . . 188
6.5 Conclusions . . . 189

7 Conclusions **191**
7.1 Summary . . . 191
7.2 Future Work . . . 194
7.2.1 New Data . . . 195
7.2.1.1 NIR and Optical . . . 195
7.2.1.2 Radio . . . 196
7.2.2 Improved Modelling . . . 197
7.2.2.1 Incorporating More Physical Processes into HOD Models . 199
7.2.3 Non-Catalogue Information . . . 200
7.2.4 Going from a Static to a Dynamic View of Halo Occupation . . . 202

Bibliography **203**

List of Figures

1.1 Figure from Betoule et al. (2014). The x-axis shows redshift, the y-axis the 'distance modulus' (a distance proxy). Results from 4 different samples of supernovae are shown, SDSS, the Supernovae Legacy Survey, and HST. The black curve is the best fit, corresponding to ΛCDM. 10

1.2 CMB temperature harmonics from the Planck Mission, Planck Collaboration, (2016). The location of the 'peaks' in this characterisation of the anisotropy gives information about the scale of acoustic oscillations in the early Universe. 13

1.3 Illustration of the evolution of a spherical over-density from Eisenstein et al. (2007). Each figure shows increasing time/decreasing redshift. The curves within each figure correspond to a different component (dark matter, gas, photons and neutrinos). The x-axis is radial distance from the original over density, the y-axis is density. The top two plots show the wave of the photon-gas plasma expanding from the over-density. The central two plots show the photons uncoupling from the gas and continuing to expand. The final two plots show gravity realigning the dark matter and gas distributions. 15

1.4 Measurements of the BAO in the SDSS BOSS survey from Anderson et al. (2014). The left two plots show the peak in the two point correlation function, and the right two plots show the corresponding sinusoid in the power spectrum/Fourier space (these plots are different representations of the same data). The top two plots show the BAO measured 'as is', the bottom two show it after the 'reconstruction' process, which corrects for small non-linearities, for more accurate measurements. 16

1.5 Illustration of weak gravitational lensing from Ellis (2010). White markers show the expected shear imparted on background galaxies (e.g. a source would be stretched/elongated along the length of the line). The blue shading represents the foreground matter distribution, from densest (white), to voids (black). 18

1.6 Figure from Suzuki et al. (2012) showing constraints on $\Omega_{M,0}$ (x-axis), and $\Omega_{\Lambda,0}$ (y-axis). Blue shading shows constraints from supernovae, green shows constraints from the BAO, orange shows constraints from the CMB, and the central small white/grey the constraints from the three simultaneously. The black line marks the space corresponding to a flat Universe (above the line corresponds to positive curvature, below negative curvature). The top left shaded area corresponds to a Universe that has always existed, with no Big Bang. It is evident that having multiple probes greatly helps in 'honing' in on one part of parameter space, corresponding to our familiar model of ΛCDM, $\Omega_{M,0} \approx 0.3$, $\Omega_{\Lambda,0} \approx 0.7$. 19

1.7 Composite image of the bullet cluster. The orange/white light is optical imaging from the Hubble Space Telescope, showing the galaxies, both in the cluster itself, and in the background. The red colouring is X-ray intensity from the Chandra space telescope, mapping hot gas. The blue shading is gravitational lensing potential, tracing where most of the mass is. The fact that the visible mass and total mass are misaligned is taken as a signal of the existence of dark matter. Image credit: NASA Astronomy Picture of the Day 24th August 2006. 21

1.8 Simulated illustration of how structure would form differently if dark matter was cold (top row), warm (middle) and hot (bottom) respectively. Pink corresponds to dense regions, blue to under-dense regions. The right hand figures show 'zoomed in' regions of the left hand figures. Image credit, ITP, University of Zurich. 23

1.9 The model matter power spectrum from linear theory (black line), with non-linear corrections (red dashed line, see section 1.2.2). Note the rough power law behaviour either side of $k_{\rm eq}$ The oscillations on small scales are baryonic acoustic oscillations, section 1.1.4.3. Figure from Daniel Baumann Part III Cambridge Cosmology notes. 26

1.10 Cartoon from Humboldt State Geospatial Online showing which parts of the electromagnetic spectrum the atmosphere is transparent to. 36

1.11 Example galaxy spectrum showing the two thermal peaks of stellar light and dust thermal emission. The black points are observations, the red curves are dust models, the blue curves are stellar thermal emission before dust absorption, the teal is stellar emission after dust, note the particular drop after absorption in the UV at $\sim 0.1\mu$m. Figure from Hayward & Smith (2014). 37

1.12 The stellar mass function at $z \sim 0.1$ from McGee et al. (2014b). 40

1.13 The stellar masses and specific star formation rates of galaxies in the local Universe in SDSS, and at low/intermediate redshifts in AGES. The dashed line shows the division between the star forming main sequence and passive galaxies. Plot from Choi et al. (2014). 41

1.14 Cosmic star formation as a function of redshift. Plot from Madau & Dickinson (2014). 42

1.15 Figure from the Hyper Suprime-Cam strategic survey proposal (Miyazaki et al., 2013). Plotted are the areas (in $\deg^2$, x-axis) and depths (AB magnitude, y-axis) of several optical extragalactic surveys. Shown in black are completed surveys, in purple ongoing surveys, in red Hyper-Suprime Cam surveys and in blue future planned surveys (all as of 2012 when the plot was made). Surveys in the top left are extremely deep, but narrow surveys; those in the bottom right are very wide (whole sky $\approx 10^{4.6}$ $\deg^2$), but very shallow. In general, contemporaneous surveys lie on some diagonal - as time goes on and instrumentation improves, the curve of accessible parameter space moves towards the top right. Exceptions can exist e.g. a survey might only be able to achieve a lower depth and area than might be expected if it is also required to have higher resolution. This figure shows optical surveys, but the behaviour of how survey power increases over time is qualitatively the same at other wavelengths. 45

2.1 Visualisation of the Fingers-of-God effect in the Sloan Digital Sky Survey from SubbaRao et al. (2008). The observer is in the centre at the bottom of each figure. The left half of the figure shows large-scale structure converting redshifts to distances ignoring peculiar motion. The right half of the figure shows the same data with redshift-space distortions 'corrected', using an approach from Tegmark et al. (2004). 50

2.2 Toy-model illustration of redshift-space distortion of the cross correlation of massive to less massive galaxies from Croft (2013). Top left: contours of an isotropic spatial correlation function. Top right: gravitational redshift added. Bottom left: peculiar motion added. Bottom right: gravitational redshift and peculiar motion added. Note that the size of the gravitational redshift has been multiplied by 500 for it to be visible - it is a much smaller effect than peculiar velocities. 51

2.3 Illustration of the same galaxy 2-point statistics through a) the spatial correlation function (first row), b) projected correlation function (second row), c) angular correlation function (third row) and d) power spectrum (bottom row). In each sub-plot the full line shows the full function, dotted line the 2-halo term, dashed line the central-satellite part of the 1-halo term, and dot-dashed line the satellite-satellite part of the 1-halo term. The 2-halo term goes to zero on small scales because we do not allow haloes to overlap, pure linear theory would extend it to arbitrarily small scales. The model correlation function here is from a Zheng et al. (2005) HOD model (see section 2.3, for $z = 1$, and $M_{min} = 10^{12} M_{\odot}$, $M_1 = 10^{13} M_{\odot}$, $\alpha = 1$, $\sigma = 0.2$ and $M_0 = 10^8 M_{\odot}$). 54

2.4 A model angular correlation function, and the impact of the integral constraint on the measured angular correlation function when it is measured in three square fields of different sizes. The angular correlation function can be measured accurately on scales much smaller than the size of the field, but is heavily biased downwards on scales comparable to the field size. The model correlation function here is from Zheng et al. (2005) (see section 2.3, for $z = 1$, and $M_{min} = 10^{12} M_{\odot}$, $M_1 = 10^{13} M_{\odot}$, $\alpha = 1$, $\sigma = 0.2$ and $M_0 = 10^8 M_{\odot}$.) 56

2.5 Illustration of the agreement of the binning approach and the non-parametric approach to correlation function calculation for a sample with $1.25 < z < 1.5$, $M_{\star} > 10^{10.6} M_{\odot}$. The error bars (the secondary lines in the case of the kernel method) represent the 16th and 84th percentiles from bootstrapping. Full details of the measurements can be found in Chapter 3. 60

2.6 Illustration of how the bias of halos of a given mass evolves with redshift. The dashed line shows $b = 1$. Halo bias increases rapidly with halo mass and redshift. By $z = 0$, bias is close to one, and slightly less than one for lower mass haloes. Halo bias taken from Tinker et al. (2010). 63

2.7 The halo mass function (comoving number density per unit of mass) as a function of halo mass, for a range of redshifts. The HMF increases with time (/decreases with redshift), and decreases with increasing halo mass. The characteristic power law with high mass exponential cut-off behaviour discussed in section 1.2.2.2 is evident. By $z = 0$ the HMF is essentially a power law over all halo masses of interest. Halo mass function taken from Tinker et al. (2010). 64

2.8 Illustration of how the spatial correlation function (left column) and galaxy comoving number density (right column) of galaxies varies with each individual Zheng et al. (2005) HOD parameter. Each model has $M_{\rm min} = 10^{12} M_{\odot}$, $M_1 = 10^{13} M_{\odot}$, $\alpha = 1$, $\sigma = 0.2$ and $M_0 = 10^{8} M_{\odot}$, apart from the individual parameter being varied, and is for $z = 1$. Blue to red corresponds to increasing parameter value. 71

3.1 The mass and redshift of galaxies, after application of the magnitude cut, star exclusion and mask, are shown here in blue. The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of $K_s < 23.5$, and the green curve the implied 90% stellar mass completeness limit, following the approach of Johnston et al. (2015). The red boxes illustrate the redshift and stellar mass selected sub-samples that I consider in subsequent sections. 79

3.2 Colour-colour diagram of all the K$<$ 23.5 sources in my sample, with stellar locus (red dotted line), and stellar cut (green dashed line) shown. 80

3.3 The redshift distributions used in the analysis for each redshift bin (arbitrary normalisation). 81

3.4 The angular correlation function for different redshift ranges and masses denoted in each panel. The lower x-axis denotes angular scale, the upper x-axis the corresponding projected comoving distance, and the y-axis the correlation function. The fainter, dashed, upper and lower bands represent the error bars on the measurements, discussed in section 2.1.2.3. For the clarity of the plot, I only show alternate stellar mass samples in the first three redshift bin subplots. 84

3.5 The measured correlation functions and the corresponding HOD best fits. Sub-figures in the same column have the same redshift, sub-figures in the same row have the same stellar mass range. The coloured filled lines are the data (blue to red corresponding to increasing stellar mass), and the lower and upper bands are 16th and 84th percentiles from the bootstrapping. The dashed black line is the model correlation function from the best fits. χ^2 values for each fit are shown in the upper right of each sub-figure. 86

3.6 Evolution of the HOD parameter $M_{\rm min}$ and M_1 as a function of stellar mass in the four redshift bins denoted in the legend. 87

3.7 Evolution of the ratio of the HOD parameters M_1 to $M_{\rm min}$ as a function of stellar mass in the four redshift bins denoted in the legend. 87

3.8 Evolution of the HOD parameter α as a function of stellar mass in the four redshift bins denoted in the legend. The dotted line represents $\alpha = 1$. . . . 88

3.9 Evolution of the HOD parameter σ as a function of stellar mass in the four redshift bins denoted in the legend 88

3.10 Plot comparing HOD parameters $M_{\rm min}$ and M_0 in the four redshift bins denoted in the legend. A one to one line is over-plotted as a guide. 89

3.11 Comparison of my HOD results with similar modelling of UltraVISTA data in McCracken et al. (2015). 90

3.12 Galaxy bias as a function of stellar mass at the four redshift bins denoted in the legend. The bias increases with stellar mass, and also increases with redshift. 95

3.13 The satellite fraction as a function of stellar mass at the four redshift bins denoted in the legend. The satellite fraction is flat, tailing off at high masses, with more satellites at low z as expected. 95

3.14 Evolution of r_0 with stellar mass. Colour represents redshift bin as denoted in the legend. 96

3.15 The stellar-mass halo ratio for different redshifts, divided into contribution from central galaxy and satellite galaxy. I plot for comparison $M_\star/M_{\rm halo} = (\Omega_b/\Omega_{\rm DM})$, and $M_\star/M_{\rm halo} = (\Omega_\star/\Omega_{\rm DM})$ from Fukugita & Peebles (2004) . . 98

3.16 The redshift dependence of the peak of the SMHR from the data compared with several values from the literature. 99

4.1 The mass and redshift of galaxies, considered after application of the magnitude cut, star exclusion and mask, are shown here in blue. The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of $K_s < 23.5$, and the green curve the implied 90 per cent stellar mass completeness limit, following the approach of Johnston et al. (2015). The red boxes illustrate the redshift and stellar mass selected sub-samples that I consider in section 4.4, with the left plot corresponding to the passive sample, and the right hand plot the star-forming sample. . . 105

4.2 The mass and sSFR of galaxies in the final sample, in the four redshifts considered in section 4.4. The blue lines delineate the various samples. The initials indicate which section corresponds to each sample: 'P'=passive, 'S'=star forming, 'H'=high mass, 'L'=low mass. 106

4.3 The cross-correlation function signal between low mass ($10.4 < \log_{10}(M_{min}/M_{\odot}) < 10.9$) and high mass ($10.9 < \log_{10}(M_{min}/M_{\odot}) < 11.4$) galaxies in VIDEO in four redshift bins. The low mass sample is selected to be passive ($\log_{10} sSFR < -11$) for the red curves, and star forming ($\log_{10} sSFR > -11$) for the blue curves. At the highest redshift, the two curves are very similar; the role of environment in determining star formation rate seems to be small. At $1 < z < 1.5$ the passive low-mass galaxies are more clustered around the massive galaxies, but only on small scales. In the $0.75 < z < 1$ and $0.4 < z < 0.75$ bins, the enhancement has reached the larger scales, and is substantial on all scales. 113

4.4 The cross-correlation function signal between low mass ($10.4 < \log_{10}(M_{min}/M_{\odot}) < 10.90$) and high mass ($10.9 < \log_{10}(M_{min}/M_{\odot}) < 11.4$) galaxies in VIDEO at two redshifts. Red denotes when the low mass galaxy is passive ($\log_{10} sSFR < -11$), blue denotes starforming ($\log_{10} sSFR > -11$). In the left two plots the massive galaxy is passive, and the right two the massive galaxy is starforming (with criterion $\log_{10} sSFR \lessgtr -11$). 115

4.5 Illustration of the three possible ways for quenching to have spatial dependence within a halo. In each figure, the pale green symbolises the extent of the dark matter halo. The black star represents the centre of the halo. The large red circle represents a high mass galaxy, that may or may not be affecting the star formation rate of lower mass galaxies, denoted by the smaller dots. The shaded pale red area denotes 'quenched areas'. The lower mass galaxies are passive (denoted by being red) when in the quenched area, and star forming (denoted by being blue) when not. In Scenario 1, galaxies are preferentially quenched when near the massive galaxy, and the massive galaxy traces the halo profile. In Scenario 2, quenching is associated with proximity to the centre of the halo, and again all galaxies trace the dark matter profile. In Scenario 3, the massive galaxy is located at the centre of the halo, and quenching is associated with proximity to the centre of the halo, which coincides with the massive galaxy. 118

4.6 An illustration of the consequences of different quenching models within my formalism. First row: probability of being quenched as a function of spatial separation. Second row: profile of galaxies within a halo. Third row: auto correlation of galaxy samples. Fourth row: cross correlation of low mass to high mass galaxy samples. Full lines correspond to Scenario 1, dashed lines to Scenario 2, dotted lines to Scenario 3. All functions (described in table 4.2) normalised, red corresponds to passive and blue to star forming where appropriate. The differences between the Scenarios are clearly much more prominent in the cross-correlations than the auto-correlations for all models. 122

4.7 The 'environment tracer' measurements and best fit model at my four redshift ranges. The solid lines are the data; the cross correlations of massive to less massive galaxies, red when the low mass galaxy is passive and blue when star forming. The error bars are the 16th and 84th percentiles of the estimator. The dashed lines correspond to the best fit model model using the quenching modification to the 1-halo term. 129

4.8 Corner plot of the posterior for the MCMC parameter estimation for the $0.75 < z < 1$ quenching model. 130

4.9 The quenching probabilities implied by the MCMC parameter fitting for the four redshift ranges is shown in red, i.e. the probability of the lower mass galaxies being quenched as a function of radial location in halo. The full thick and two thin lines are the pointwise median and 16th and 84th percentiles respectively of the sampled models. The dashed red lines are the model corresponding to the best fit parameters. The horizontal black dashed lines are global passive fraction for the lower mass galaxies at that redshift. . . . 131

4.10 The halo mass of the environment probed at each redshift I make measurements for. Error bars are the 16th and 84th percentiles of the MCMC fitting samples. 131

4.11 The cross biases of each of my correlation functions. The cross bias is the geometric mean of the bias of each subsample used in the cross-correlation. 132

4.12 Estimate of the probability of being quenched as a function of radial location from my best fit models when the central is passive (red), and when the central is star forming (blue). Full lines correspond to the pointwise 16th, 50th and 84th percentiles, dashed line to the quenching model of the best fitting parameters. 133

5.1 The geometry of the UDS and UltraVISTA fields. The red points are the galaxy locations from Bowler et al. (2015). The blue points are the random points chosen to cover the fields used for the construction of the random galaxy catalogue for the calculation of the correlation function. The three galaxies in the UltraVISTA field that are not surrounded by blue points are the $z = 6$ sources detected in the deep (as opposed to ultra-deep) part of the UltraVISTA field, that I do not include in this study. The overall shape of the fields is predominantly determined by the part of the sky that the multiple different surveys overlap in. The small scale gaps and holes are foreground stars and detector artefacts etc. See figures 1 and 2 of Bowler et al. (2014) to see how the irregular footprints arise from the intersection of the sky patches covered by different surveys. 150

5.2 The angular correlation function for the sample of bright ($M_{UV} < -21.125$) $z \sim 6$ LBGs from Bowler et al., (2015). The figure shows the correlation function estimated both with a binning method (blue points), and a kernel smoothing method (red curve, with dotted lines showing uncertainty). Where the binned correlation function dips to negative values corresponds to where there were no galaxy pairs in the bin. 154

5.3 The correlation coefficients (covariance normalised) of my measurements. Blue values are positive correlations, red values are negative values. The blue diagonal corresponds to the standard deviation measurements (a random variable is always perfectly correlated with itself). 155

5.4 Comparison of the measurements of the correlation function (red curve and shaded area) with the five different models I fit. The two blue curves correspond to models with only M_{min} free, the dashed curve has DC = 1 and the full line DC = 0.6. On linear scales all models are very similar, apart from the bias only model, as the number density constraint is restricting the model from going too high. Only the M_1 free model allows the small-scale amplitude to vary independently. 157

5.5 Triangle plot of the posterior from the MCMC fitting for the HOD model with M_{min}, M_1 and DC free (masses in log base ten Solar mass units). Dashed lines on the one dimensional single parameter plots are 16th, 50th and 84th percentiles. 158

5.6 Comparison of the LBG correlation function (red curve, 1-σ uncertainties in the lighter curves), with the lower luminosity Harikane et al., (2016) measurements, and the dark matter angular correlation function (black curve). 161

5.7 Comparison of the results from the LBG sample in this chapter, with comparable measurements of lower luminosity LBGs from Harikane et al. (2016). Top: M_{min} as a function of absolute UV luminosity threshold (in units of Solar mass). Bottom: galaxy bias as a function of absolute UV luminosity threshold. The results from the six different models are shown for comparison (x-axis values slightly offset for each model for clarity). 162

5.8 Luminosity (monochromatic luminosity at 1500$\mathring{A}$; that is to say, λf_λ where λ is the wavelength and f_λ is the flux per wavelength) to halo mass ratios as a function of halo mass, derived using abundance matching with the Bowler et al., (2015) luminosity functions and Behroozi et al., (2013) halo mass functions. The results are shown for both Schechter function fit and the double power law fit, for $z = 5$, $z = 6$ and $z = 7$. In bold are the best fitting functions, double power law for $z = 6$ and $z = 7$, Schechter for $z = 5$, and the dashed line shows the alternative. 172

6.1 Stellar masses and redshifts (blue points) of all $K < 23.5$ galaxies in the Laigle et al. Horizon-AGN mock catalogue. The red lines denote the bins used in Chapter 3. The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of $K_s < 23.5$, and the green curve the implied 90% stellar mass completeness limit, following the approach of Johnston et al. (2015). See figure 4.1 for the corresponding figure in the VIDEO observations. 181

6.2 The angular correlation function for each stellar mass and redshift bin, measured in the Horizon-AGN mock catalogue. 182

6.3 Comparison of the VIDEO and Horizon-AGN clustering measurements for the stellar mass and redshift bins. The thin lines are the errors on the ratio, with the errors on both the VIDEO and Horizon-AGN measurements propagated through.The dashed line corresponds to $\omega_{HORIZON} = \omega_{VIDEO}$. 183

6.4 Best fit M_{min} and M_1 for the Horizon-AGN clustering measurements, as a function of galaxy sample threshold stellar mass. Equivalent measurements in VIDEO from Chapter 3 shown for comparison. 184

6.5 Comparison of the galaxy masses direct from the simulation, and as inferred from fitting to the simulated photometry for a range of redshifts. Dashed lines correspond to offsets of 0.2 dex, grey points correspond to the median. Plot from Laigle et al., in prep. 186

6.6 Comparison of the central stellar mass to halo mass relation directly from Horizon-AGN (red curve), and from observations (dark shading, from an abundance matching analysis, Moster et al., 2013). Over-plotted are the results from the clustering analysis (green triangles from VIDEO, blue triangles from Horizon-AGN). Teal lines show constant star formation efficiency (of 0.1 and 0.7 respectively). Figure courtesy of Y. Dubois, with data points from this chapter overplotted. 188

7.1 Schematic of how the multi-wavelength data over the VIDEO fields will evolve over the coming years. The rectangles are the VISTA paw-prints, teal corresponding to VIDEO, pink to the new VEILS survey. The white circles have deep DES coverage, and the yellow circles are MIGHTEE. Figure from the MIGHTEE proposal (M.Jarvis, private communication) 195

7.2 Preliminary results of developing methods to measure unresolved flux in MIGHTEE. The x-axis shows inverse angle/angular frequency, and y-axis the power spectrum of the flux. The green line (binned into the red points and error bars) shows the power spectrum of the unresolved flux. The blue line shows the predicted power spectrum from synchrotron radiation from our own Galaxy from Santos et al. (2005). We would anticipate that we would be unable to measure the clustering of the unresolved sources on scales where Galactic synchrotron dominates. The black dotted line shows a fit to the amplitude to the noise - the fact that the power spectrum is not just flat shows that the sources are not uniformly distributed. 198

7.3 All the measurements in Coupon et al. (2015) - stellar mass function, clustering (angular and projected) for several stellar mass ranges, and galaxy-galaxy lensing. The black curve that goes through all these points is all from just one model simultaneously fitted to all the data. 199

7.4 Cartoon of combining data from three wide deep extra-galactic surveys at one redshift (assuming that clustering continues to increase with stellar mass over the relevant mass scales). Using the Coupon et al. (2015) model, all of this data (at one redshift) could go into constraining one model, rather than a separate model at each stellar mass. 200

List of Tables

3.1 Parameters from the angular correlation function evaluation and HOD fitting, with corresponding χ^2/d.o.f. values for the fits. 85

4.1 Number of galaxies for each redshift, stellar mass, and star formation rate bin, with the associated passive fraction for the low mass galaxies. H denotes the high mass ($10.9 < \log_{10}(M_\star/\mathrm{M}_\odot) < 11.4$) galaxies, PL the passive low mass galaxies, and SL the star forming low mass galaxies ($10.4 < \log_{10}(M_\star/\mathrm{M}_\odot) < 10.9$, $\log_{10} sSFR \lesseqgtr -11$). Note that these counts are actually the sums of the weights of the galaxies in each redshift bin. 107

4.2 Summary of the resulting profiles, along with auto- and cross-correlation functions for different quenching mechanisms within a halo. δ denotes a Dirac delta, $*$ denotes a 3D convolution 121

4.3 The constraints on our quenching and bias model from our MCMC results. The errors are the 16th and 84th percentiles of the posterior. The cross biases are labeled by what pairs of samples they correspond to; H-PL corresponds to the high mass sample to low mass passive sample, H-SL the high mass to low mass star forming sample. 132

5.1 Fitted HOD parameters from the MCMC. Also shown are the corresponding satellite fractions ($f_{\rm sat}$) and galaxy biases (b) and fit reduced χ^2 of the samples. Quantities in brackets are either fixed in the model, or fixed as a function of other parameters in the model. Masses are in Solar mass units (log base ten). Note that values and error bars quoted are the 16th, 50th and 84th percentiles of the posterior, as opposed to the peak values. This makes very little difference apart from the posterior for the duty cycle value for the M_{min}, M_1 and DC free model, which is peaked at DC = 1 and hence only has one tail, see Fig. 5.5. The lower luminosity parameter values are taken directly from Harikane et al. (2016), apart from satellite fraction, which I calculate. 159

5.2 Comparison of abundance matching results to clustering fits for the $z \sim 6$ LBGs. Columns are (1) LBG sample used, (2) LBG threshold absolute magnitude, (3), observed comoving number density (Mpc^{-3}), (4) the minimum halo mass in the most straightforwards abundance matching scheme (log base ten Solar mass units), (5) the model comoving number density (Mpc^{-3}) of the best fit HOD models in this work and Harikane et al., (2016) without incorporating duty cycle, (6) the corresponding minimum halo mass from the HOD model (log base ten Solar mass units). 160

5.3 Actual and expected number of galaxies in each field. The columns are: field used, the field angular area (in deg^2), the actual number of galaxies in the field ($N_{\rm a}$), the actual angular galaxy density in the field ($\rho_{\rm a}$, in deg^{-2}), the expected number of galaxies in the field if it had the mean density ($N_{\rm e}$), the expected angular galaxy density in the field - all identical figures as I am considering deviations from the mean density ($\rho_{\rm e}$, in deg^{-2}), standard deviation from Poisson statistics, equal to the square root of $N_{\rm e}$ ($\sigma_{Poisson}$), standard deviation from cosmic variance, estimated from the clustering measurements (σ_{CV}), and the Poisson and cosmic variance errors added in quadrature (σ_{Total}). 170

Chapter 1

Introduction

This chapter reviews three major interlinking components of 21st Century astrophysics; cosmology, the large-scale structure of the Universe, and the formation and evolution of galaxies.

1.1 Cosmology

Cosmology is the study of the Universe as a whole, and seeks to understand and answer questions that have been asked throughout human history: how did the Universe come to be, how old is the Universe, how large is the entire Universe, and what is humanity's place in it?

Cosmological questions and ideas have been discussed since pre-history, from the Ancient Greek Ptolemaic model that survived well into the Medieval period, through the geocentrism versus heliocentrism debates of the 16th century, through to 19th century realisations from geology and natural history that suggested that the Earth must be hundreds of millions, if not billions of years old. Finally the modern field of cosmology is often considered to have started in the 1920's with Edwin Hubble's two revolutionary discoveries that each founded entirely new fields:

- *Extragalactic astronomy* - the discovery that nebulae were *not* part of the Milky Way, and were in fact their own 'island universes', systems of stars like the Milky Way at vast distances from us

- Observing that galaxies are all moving away from us, showing that the Universe was expanding, and that *cosmology* could be studied as a modern quantitative science

Since then the essential idea of an expanding Universe has remained, now justified through multiple different probes, and measured to great precision (albeit with several substantial changes over the decades). Current cosmology seeks to improve constraints on our model in order to rule out variants on the main theory, as well as attempting to identify the true nature of the main 'unknowns' of the model, in particular dark matter and dark energy.

We use the units in which $G = c = 1$ in this chapter.

1.1.1 Big Bang Cosmology

In this sub-section we discuss how cosmology and the concept of an expanding Universe is expressed and derived in General relativity (Einstein's theory of gravity that emerged at the beginning of the 20th Century, alongside Hubble's discoveries).

The central premise of general relativity is that gravity can be understood in terms of matter causing spacetime to curve - *"Spacetime tells matter how to move; matter tells spacetime how to curve"*[1]. This is expressed mathematically by the Einstein field equation (Einstein, 1915):

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1.1}$$

where $G_{\mu,\nu}$ is the Einstein tensor which represents curvature ($G_{\mu\nu} = R_{\mu,\nu} - \frac{1}{2}Rg_{\mu\nu}$, $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, $g_{\mu\nu}$ the metric) and $T_{\mu\nu}$ is the stress-energy tensor which represents matter (incorporating energy density, momentum density, pressure and shear stress)[2]. Λ is a constant, typically known as the Cosmological Constant, or the dark energy term, and was initially believed to be equal to zero, but is now postulated to have a small non-zero value, see section 1.1.2.

A 'metric' in general relativity describes how distances are measured in the curved

[1] John Archibald Wheeler in "Geons, Black Holes, and Quantum Foam"
[2] The sign of various terms can vary according to convention.

spacetime. Any homogeneous ('invariant under spatial translation') and isotropic ('invariance under spatial rotation') Universe in general relativity must have a metric of the form:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\right), \tag{1.2}$$

known as an Friedmann-Lemaître-Robertson-Walker (FLRW) metric (e.g. Robertson, 1935), where s is proper time/distance, t is a time coordinate, r a radial coordinate, and ϕ and θ are spherical coordinates. The parameter k is a constant, and corresponds to the spatial curvature of the Universe, where $k = 0$ is flat, $k > 0$ is elliptic and $k < 0$ is hyperbolic. Rescaling r allows k to be set to -1, 0 or +1. The variable a is a function of time and is the *scale factor* of the Universe. The scale factor describes the 'size' of the Universe at a given time, but not in absolute terms - ratios between the value of a at two different times are meaningful, but the value of a at a given time does not correspond to a particular scale.

The stress-energy tensor for perfect fluids[3] must be of the form $T = diag[\rho(t), P(t), P(t), P(t)]$ where $\rho(t)$ is density as a function of time, and $P(t)$ is pressure as a function of time. Solving the Einstein equation for that metric yields the two Friedmann equations (Friedmann, 1922):

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}(\rho + 3P) + \frac{\Lambda}{3}, \tag{1.3}$$

$$\left(\frac{\dot{a}}{a}\right)^2 \equiv H^2(t) = \frac{8\pi}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}. \tag{1.4}$$

The final piece of information required to solve the equations is how pressure behaves as a function of density for the fluid in question. The fluid equation in an expanding universe is:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0. \tag{1.5}$$

[3]No heat conduction makes the time-space components zero, no viscosity makes the off-diagonal space-space components zero.

Taking the equation of state of the fluid as $P = w\rho$, we can solve to find $\rho \propto a^{-3(1+w)}$. Given this relation, the Friedmann equations can be solved. The behaviour of the Friedmann equations for different equations of state, and different values of k are well understood. The interpretation of the constant k is that it corresponds to the 'curvature' of the Universe, with positive values corresponding to a Universe with spherical/elliptic geometry, negative values correspond to hyperbolic geometry, and $k = 0$ to a flat universe. There are three $k = 0$ solutions of particular interest to us, which we discuss below.

Firstly, a Universe filled with a pressure-less dust ($P = 0$, $w = 0$) has $\rho \propto a^{-3}$. This agrees with our intuition that if you double the length of the side of a cube, the volume goes up by $2^3 = 8$, and then density of the substance inside goes down by that amount. Secondly, a radiation filled Universe has $\rho \propto a^{-4}$ (extra factor of a for the photons losing energy through being stretched by the expansion of the Universe), corresponding to $w = 1/3$. Finally, if a substance were to have the remarkable property of having a fixed density regardless of box size ($\rho \propto a^0$ e.g. if the box were doubled in size there would now be twice as much of the substance), then $w = -1$ and $P = -\rho$ - negative pressure! This mysterious substance turns out to be equivalent (at least as far as the dynamics of an expanding Universe are concerned) to having $\Lambda \neq 0$. This means 'Λ' can be thought of either as a peculiar fluid or field filling the entire Universe (when it is normally known as 'dark energy'), or as a modification to Einsteins field equation (when it is normally known as a 'cosmological constant'). For the purposes of this thesis I use the concepts interchangeably.

In a radiation dominated Universe, one can derive $a(t) \propto t^{1/2}$, in a matter-dominated Universe $a(t) \propto t^{2/3}$, and in a dark energy dominated Universe $a(t) \propto \mathrm{e}^{Ht}$ (where H is a constant). In general, in an expanding Universe with a mix of components, the Universe will be dominated by the component which rises in density fastest as volume is decreased at early times, and by the component that dilutes slowest at late times. In an expanding Universe with a mixture of radiation, matter and dark energy (which we do indeed believe to be the case, see section 1.1.2), the Universe will be radiation dominated initially, then transition to being matter dominated, before finally transitioning to being dark energy dominated and expanding exponentially forever.

It is often convenient to rewrite the second Friedmann equation (eq. 1.4) in the form:

$$1 = \frac{8\pi}{3H^2}(\rho_M + \rho_R) - \frac{k}{H^2a^2} + \frac{\Lambda}{3H^2} = \Omega_M + \Omega_R + \Omega_k + \Omega_\Lambda, \tag{1.6}$$

where $\Omega_M = \frac{8\pi}{3H^2}\rho_M$, $\Omega_R = \frac{8\pi}{3H^2}\rho_R$, $\Omega_k = -\frac{k}{H^2a^2}$ and $\Omega_\Lambda = \frac{\Lambda}{3H^2}$; ρ_M ρ_R are the matter and radiation energy densities, the Ω_i are the fractions of the critical density (defined as $\rho = \frac{3H(z)^2}{8\pi G}$) made up by each component ($i = R$ is for radiation, $i = M$ for matter, $i = k$ for curvature, $i = \Lambda$ for dark energy). In general all these quantities are functions of time, and track how the dominant component of the Universe changes over cosmic time. We denote the value of these parameters in the Universe today as $\Omega_{i,0}$.

1.1.2 ΛCDM

As already discussed, Hubble's observations of the expanding Universe (Hubble, 1929) was the first suggestion of a non-eternal Universe. These observations, wedded with the theoretical underpinning of general relativity led to Big Bang cosmology growing in popularity among the astronomical community in the 1940's and 1950's.

One alternative interpretation to the Big Bang hypothesis was the 'Steady-state' model, in which the Universe *is* expanding, but new matter is continuously being created 'in the gaps' in such a way that the universe doesn't change over time. By the 1960's observations such as the Cosmic Microwave Background (CMB, Penzias & Wilson, 1965, see section 1.1.4.2) and radio source counts (Ryle & Clarke, 1961, see section 1.3.3) were generally accepted to have ruled out most final objections to the Big Bang hypothesis - as well as offering brand new lines of inquiry into our understanding of our Universe.

Since the late 1990's, a generally accepted Cosmological model has emerged that is flat ($k = 0$), and consisting of $\sim$ 70% dark energy/cosmological constant ('Λ') and $\sim$ 30% matter, of which $\sim$ 80% is cold dark matter ('CDM', a hypothesised as of yet unknown form of matter with very low interaction rates), and $\sim$ 20% baryonic[4] matter (protons, electrons, atoms, stars, galaxies - everything colloquially considered 'normal' matter). Early hints for the possible existence of a form of dark matter first emerged in the 1930's with measurements of galaxy velocity dispersions in clusters (Zwicky, 1937), and was consolidated

[4]In astrophysics it is common to refer to all matter that is not Dark Matter as baryons e.g. electrons are included in baryons even though they are leptons in particle physics.

by measurements of galaxy rotation curves in the 1980's (e.g. Rubin et al., 1980) - see section 1.1.5.1. Similarly, for dark energy there were early suggestions as to the possibility of the existence of a dark energy component to the Universe in the early 1990's (mainly motivated by measurements of large-scale structure and globular cluster ages e.g. Efstathiou et al., 1990; Carroll et al., 1992) - before wide-spread acceptance within the astronomical community from the supernovae observations in Riess et al. (1998) and Perlmutter et al. (1999) (see section 1.1.4.1). This model of a flat Universe with substantial dark matter and dark energy components is known as 'ΛCDM' - a startling result that everything we commonly observe in the Universe (gas, planets, stars, galaxies...) are only a small fraction of the total energy budget of the Universe.

Finally, in the 2010's we have multiple novel probes of cosmology, alongside iconic measurements like galaxy redshifts and the CMB, that all agree with this portrait of the Universe and have moved the field into a regime where we have precision, sub-percent constraints on the relevant parameters of the model (e.g. Planck Collaboration, 2016). Although there remain some inconsistencies (see Bull et al., 2016 for a discussion) this model is able to explain a huge wealth of data with only a few free parameters. Future challenges for the theory are to be able to constrain the equation of state of dark energy, the initial matter power spectrum, the nature of inflation (a hypothesised extremely brief period of exponential expansion at the very beginning of the Universe, see Liddle, 1999), and the number and mass of neutrino species (see Lesgourgues & Pastor, 2014).

1.1.3 Redshift as time and distance measure

Hubble's famous discovery of the expanding Universe hinged on observations that the redshift of a source's spectrum was proportional to its distance from the observer:

$$z \approx \frac{H_0}{c} d, \tag{1.7}$$

where z is the redshift (e.g. the spectrum is shifted by a factor of $1+z$), H_0 is Hubble's constant, and d is the distance to the galaxy. The conventional way to interpret this cosmological redshift is say that both source and observer are 'stationary' relative to 'background

space', and that the distance between them is increasing because the space between them is 'stretching'. Then the interpretation of redshift is that the light waves are stretched with the Universe as the waves pass through free space from source to observer. In reality the phenomenon is a general relativistic effect, that has to somehow be shoehorned into conventional language. However this analogy illustrates the connection between scale factor, distance and redshift (the above expression only holds locally): $a = 1/(1+z)$, where photons emitted when the scale factor was a will be observed to have redshift z today (where a is scaled to equal 1, and $z = 0$ is no redshift, looking at objects near to the Earth). In our Universe, objects in the past are observed to have positive z, which corresponds to the Universe expanding. If the Universe were collapsing, then z would be negative, as the Universe was larger in the past and everything is rushing towards us. Because our best model of the Universe has a monotonically increasing as a function of time it is meaningful to use z and a as proxies for time, and distance from observer[5].

The expressions in this sub-section are valid for $\Omega_k = 0$, which is believed to be the case for our Universe, but see Hogg (1999) or a cosmology textbook for the more general expressions.

1.1.3.1 Physical and Comoving Distance

It is convenient to define the normalised Hubble parameter:

$$E(z) \equiv \sqrt{\Omega_{\mathrm{M},0}(1+z)^3 + \Omega_{\Lambda,0}} \tag{1.8}$$

e.g. $H(z) = H_0 E(z)$, where H_0 is Hubble's constant, the Hubble parameter in the Universe today.

The line of sight comoving distance D_C from the observer to a source of redshift z is defined as:

$$D_C = \frac{c}{H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')} \tag{1.9}$$

[5]This is not the case in general; in universes that expand, and then collapse, for an observer during the collapse some values of a correspond to multiple points in time.

In a flat Universe, the transverse comoving distance for two bodies both at the same redshift, at angular separation $\delta\theta$, is simply $D_C\delta\theta$, although it is a more complex expression in a non-flat Universe.

1.1.3.2 Angular and Luminosity Distance

The angular diameter distance D_A from the observer to the source is defined to be the ratio between the physical size of an object and its angular size in radians. It is related to the comoving distance by:

$$D_A = \frac{D_C}{1+z}. \tag{1.10}$$

The angular distance has the curious property in some universes (including our own) that it reaches a maximum as a function of redshift, and then starts decreasing at higher redshift. This corresponds to an object getting smaller as it recedes in distance (and time), but eventually starting to get larger again after a certain distance. This happens because the expansion of space 'blows up' the path the lights rays travel and expands them, so that very small objects in the early Universe can appear very large.

The luminosity distance D_L from the observer to the source is defined to satisfy:

$$F = \frac{L}{4\pi D_L^2}. \tag{1.11}$$

where F is the flux measured by the observer, and L is the source luminosity. It is related to the comoving distance by:

$$D_L = D_C(1+z) \tag{1.12}$$

1.1.3.3 Look-Back Time

Objects at a redshift z exist a time t_L ago, where

$$t_L = \frac{1}{H_0}\int_0^z \frac{\mathrm{d}z'}{(1+z')E(z')}. \tag{1.13}$$

Integrating the expression to infinity ($z = \infty$) gives the total age of the Universe.

1.1.4 Cosmological Probes

Since Hubble's discovery of the expanding Universe, astronomers have found several other ways of learning about the structure and history of the Universe - we discuss some of the key methods below.

1.1.4.1 Supernovae and Standard Candles

If you have confidence in your cosmological model, you can convert redshifts into distances and look-back times as per section 1.1.3. However if you want to *learn* about cosmology, it is necessary to measure both the redshift and the distance to sources independently. Unfortunately measuring the distance to astronomical objects is difficult. One key approach is the use of 'standard candles'. Given a source where the intrinsic luminosity is known, and the flux from the object received on Earth measured, application of an inverse square law (using the luminosity distance) can give the distance to the source (an object will be fainter when observed from Earth than a closer object of identical intrinsic brightness). This reduces the problem to finding astronomical objects where the intrinsic luminosity is known. These are known as standard candles (or sometimes 'standardisable' candles, as often the population will not all have the exact same luminosity, but will fall on some scaling relation).

The standard candles most applicable to cosmology are Type Ia supernovae (generally accepted to be exploding white dwarfs), which are believed to have a relatively small range of luminosities. White dwarfs below the Chandrasekhar mass, $\sim 1.4 M_{\odot}$, are believed to be relatively stable. However above this mass white dwarfs are unstable, and typically explode as supernovae. Although the exact mechanism is not perfectly understood, most models and observations suggest that Type Ia supernovae are caused when a sub Chandrasekhar mass white dwarf slowly accretes matter, and then explodes when it reaches near the critical mass, meaning most Type Ia's are from bodies with similar masses giving them similar luminosities[6]. See Baade & Zwicky (1938), Kowal (1968), Tammann (1979) and Branch

[6]Type Ia Supernovae luminosities are also observed to have moderate dependence on the width of their

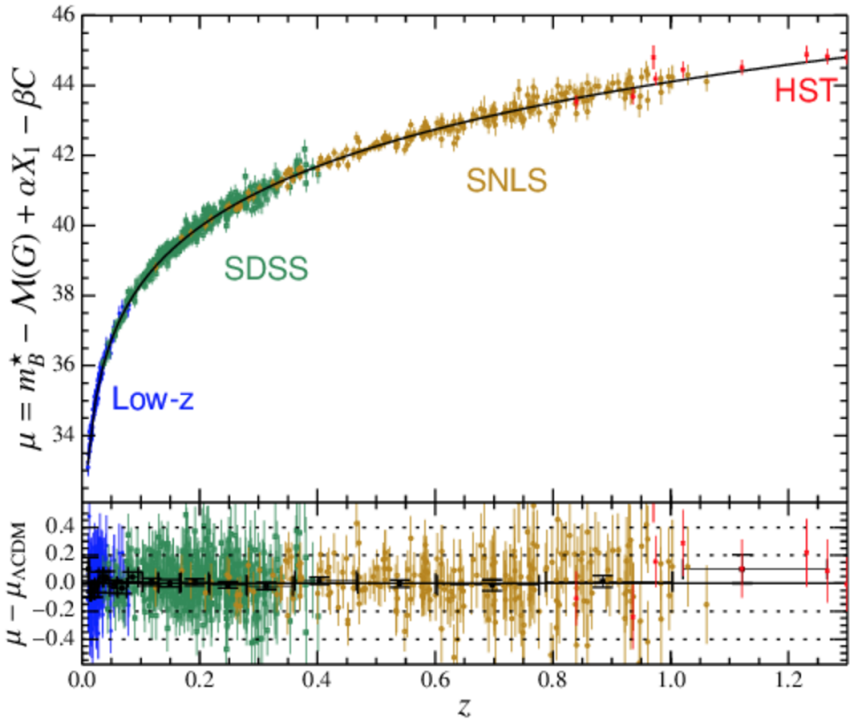


Figure 1.1: Figure from Betoule et al. (2014). The x-axis shows redshift, the y-axis the 'distance modulus' (a distance proxy). Results from 4 different samples of supernovae are shown, SDSS, the Supernovae Legacy Survey, and HST. The black curve is the best fit, corresponding to ΛCDM.

(1998) for a sample from the literature of how our understanding of supernovae as standard candles (and the corresponding cosmological applications) evolved historically, and see Hillebrandt & Niemeyer (2000) for a review of possible Type Ia supernovae mechanisms.

Supernovae light curves are now measured for supernovae up to very high redshift, and provide the clearest evidence that our Universe has a non-negligible dark energy component (Riess et al., 1998; Perlmutter et al., 1999) - see Figure 1.1.

Typically standard candles (and distance measures in general) need some calibration, and our distance measures on large scales sit atop a range of interconnected calibrated measures of distance, known as *the cosmic distance ladder*. For example, Cepheid variable stars (which are also standard candles) are too faint to see far beyond the Milky Way, and supernovae are too infrequent to be used to determine the distance within the Milky Way. However we can use nearby galaxies observed with both supernovae *and* Cepheid variables

light curve e.g. in order to get the precision necessary for cosmology, it is necessary to not only track the peak flux, but also how long it stays bright, e.g. Firth et al. (2014).

to calibrate supernovae luminosities. Cepheid variable stars might then be calibrated from parallax distances or another distance measure.

Other phenomena that can act as standard candles on cosmological scales include the Tully-Fisher relation (Tully & Fisher, 1977) and active galactic nuclei (AGN, see section 1.3.6) dust time ages (Oknyanskij, 1999). The Tully-Fisher relation is an observed power law scaling relation between absolute galaxy luminosity (or stellar mass) and rotational velocity for spiral galaxies (McGaugh, 2012; Tiley et al., 2016) - thus by measuring the rotational velocity, one gets direct information about the absolute magnitude (Saintonge et al., 2008). With regards to AGN dust lags, AGN have variability on human time frames, and the lag between an increase in optical brightness and the dust/infrared brightness scales with absolute luminosity (as the lag is related to the physical size of the object, due to an increased radius that occurs due to the sublimation radius of the dust). By measuring the lag, we can measure the absolute luminosity of the source, and use it as a standard candle (Hönig et al., 2017). Other groups have suggested using the AGN broad line region rather than dust lags (e.g. Watson et al., 2011).

1.1.4.2 Cosmic Microwave Background

One of the most iconic and precise cosmological probes is the Cosmic Microwave Background (CMB). The CMB is a 'bath' of photons that permeates the entire Universe, which we can observe as a background of microwave radiation present in every direction we look in the sky.

Shortly after the Big Bang the Universe was filled with a coupled plasma of photons, protons and electrons. This prevented the photons from free streaming as they were constantly scattering off electrons. However eventually the Universe cooled sufficiently that the electrons and protons combined to form neutral atoms ('recombination'). The interaction rate between the photons and the baryons reduced dramatically as the number density of free electrons dropped dramatically, and the mean free path of photons became larger than the Hubble horizon at the time. These photons are effectively a photon gas that fills the Universe, and cools as the Universe expands, and eventually some of these photons hit the Earth, which we observe at the CMB. The existence of the CMB was first hypothesized

in Alpher et al. (1948), and first detected by Penzias & Wilson (1965), see Scott & Smoot (2006) for a more recent review of the theory.

The CMB is black body radiation corresponding to 2.73K, peaking in the microwave at 160.23 GHz. This corresponds to a radiation temperature of around 3000K at recombination. The very existence of this afterglow, and the measurement of this temperature, is often taken as evidence of the Big Bang (alternative theories to the Big Bang would have to find another mechanism to produce this background). Because the Earth's atmosphere is near opaque at the wavelengths where the CMB spectrum peaks, the earliest CMB experiments were on balloon experiments (e.g. Benot, 2002), with the most powerful measurements now from space telescopes (COBE, Smoot, 1999; WMAP, Hinshaw et al., 2013; Planck, Planck Collaboration, 2016), although now ground based telescopes have also been able to give valuable high resolution polarisation information (e.g. the Atacama Cosmology Telescope, Thornton et al., 2016).

The real power of the CMB however is in its anisotropy - it has tiny $\sim$1 in 10,000 temperature fluctuations which contain a wealth of information about the structure of the early Universe. The CMB anisotropy is conventionally measured by expanding temperature maps in spherical harmonics (see Figure 1.2 - these can be related to power spectra and correlation functions, see section 2.1.1). Modelling these spherical harmonics can give very precise constraints on our cosmological models. The anisotropy has two causes a) anisotropy from the pattern of under and over densities at the time of recombination ('primary anisotropy') and b) things that have 'happened' to the CMB photons as they free-streamed since then ('late time anisotropy').

The spherical harmonics can be extremely well modelled and give very accurate cosmological constraints. The harmonics have several peaks in the range $l \sim 50 - 1000$, and behaves like a power law for large l values. CMB measurements give comparatively little information on dark energy as it was a subdominant component of the Universe at the early times when the CMB was formed. However dark energy does still have a small effect on CMB harmonics through the Integrated Sachs Wolfe (ISW) effect (Sachs & Wolfe, 1967). In general as the CMB photons free-stream towards us they can 'gain energy' and are blueshifted when they enter potential wells, and are then redshifted when they expend

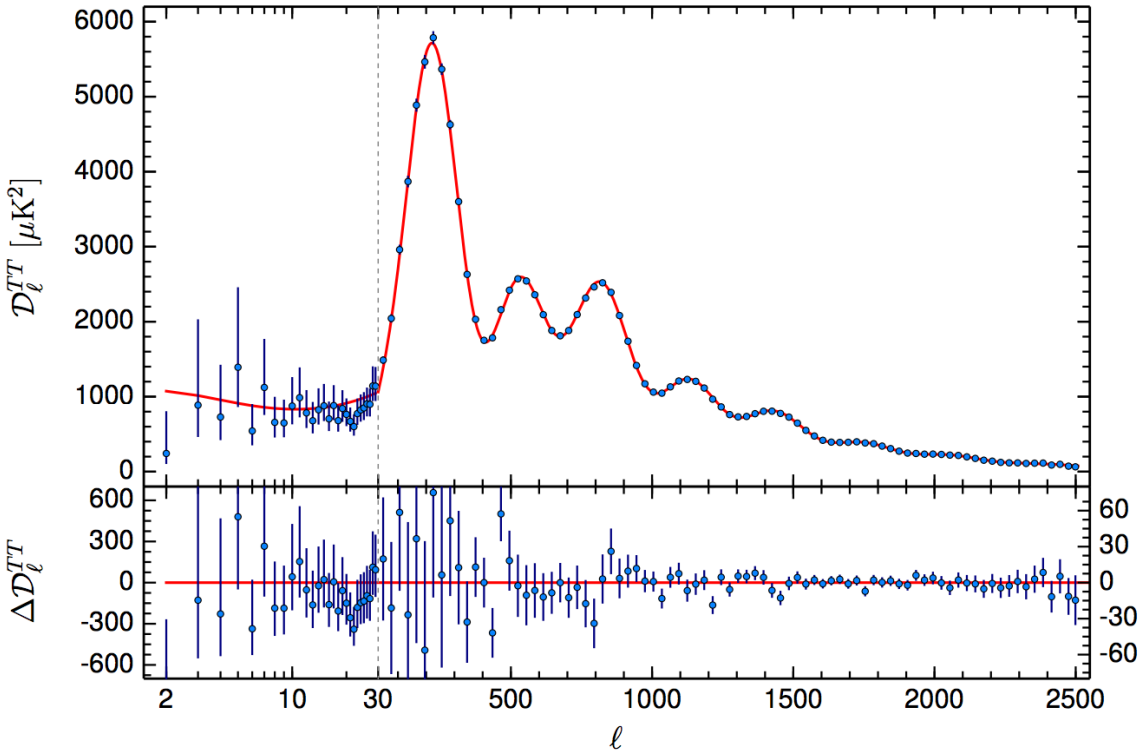


Figure 1.2: CMB temperature harmonics from the Planck Mission, Planck Collaboration, (2016). The location of the 'peaks' in this characterisation of the anisotropy gives information about the scale of acoustic oscillations in the early Universe.

energy leaving these wells. The opposite also happens with photons entering voids. In a matter dominated Universe, for linear perturbations these effects cancel out, and there is no net effect on the photons as they travel through over- and under-densities. However in a dark energy dominated Universe the over-densities 'decay' slightly over the time period that the photons are inside the potential well, meaning they need to expend less energy to escape (see Dupé et al., 2011 for a review of the ISW effect).

Another important late time anisotropy is the Sunyaev-Zel'dovich effect (Sunyaev & Zeldovich, 1970). This phenomenon occurs when CMB photons pass through the hot gas in a massive cluster/massive dark matter halo. The photons are inverse Compton scattered off the energetic electrons in the hot gas plasma, giving them an energy boost, which we view as a 'hot patch' on the CMB. This can be linked to estimates of the dark matter halo mass etc. (see Desert, 2000 for a review).

The theory of the CMB goes far beyond that discussed here. For example, the CMB

is also polarised, and measuring the E and B modes (effectively the div and curl of the polarisation field) can give valuable information about a range of processes, including potentially primordial gravitational waves (Antolini et al., 2013). There is also hypothesised to be a cosmic neutrino background, with smaller anisotropies (as neutrinos decoupled earlier than photons), but detecting this background is believed to be challenging with current technology, see Faessler et al. (2017) for a discussion of some possible approaches.

1.1.4.3 Baryonic Acoustic Oscillations and Standard Rulers

Analogous to a standard candle, 'standard rulers' can give independent distance measures, which when combined with redshift measurements give cosmological information. Standard rulers are bodies of known physical size, where the angular size can be measured - comparing the two gives the distance to the source. Standard candles give the luminosity distance, and standard rulers give the angular distances (see section 1.1.3), so the two approaches are complementary.

In the early Universe, photons and baryons (in the form of a hot gas) are tightly coupled through Thompson scattering. Dark matter is only faintly coupled through gravity. Gravity and pressure create sound waves in the photon-gas soup. If one spherical over-density is considered, it will propagate outwards at a speed $c_s = c/\sqrt{3(1+R)}$ where $R = 3\rho_b/4\rho_\gamma = \Omega_{b,0}/(1+z)$. Eventually, the Universe has expanded enough that the mean free path of the hot gas interacting with the photons is greater than the Hubble scale, and the baryons decouple from the photons. The photons free stream and become the cosmic microwave background, see section 1.1.4.2. The baryons remain at that fixed distance from the original over density, leaving a signal of matter having a greater likelihood of being separated at that scale. Over subsequent time, gravity causes the baryonic and dark matter to realign. Finally galaxies form from the baryonic matter. See Figure 1.3 for a diagram of how the distributions of the different components evolves over time. By measuring the large scale distribution of matter (via the two point correlation function, see section 2.1.1), one can detect this scale, by noting an excess of galaxy pairs at this angular separation, see Figure 1.4. The length scale of these 'baryonic acoustic oscillations' (BAO) behave as a standard ruler. These measurements can also be performed in Fourier space, where the BAO scale

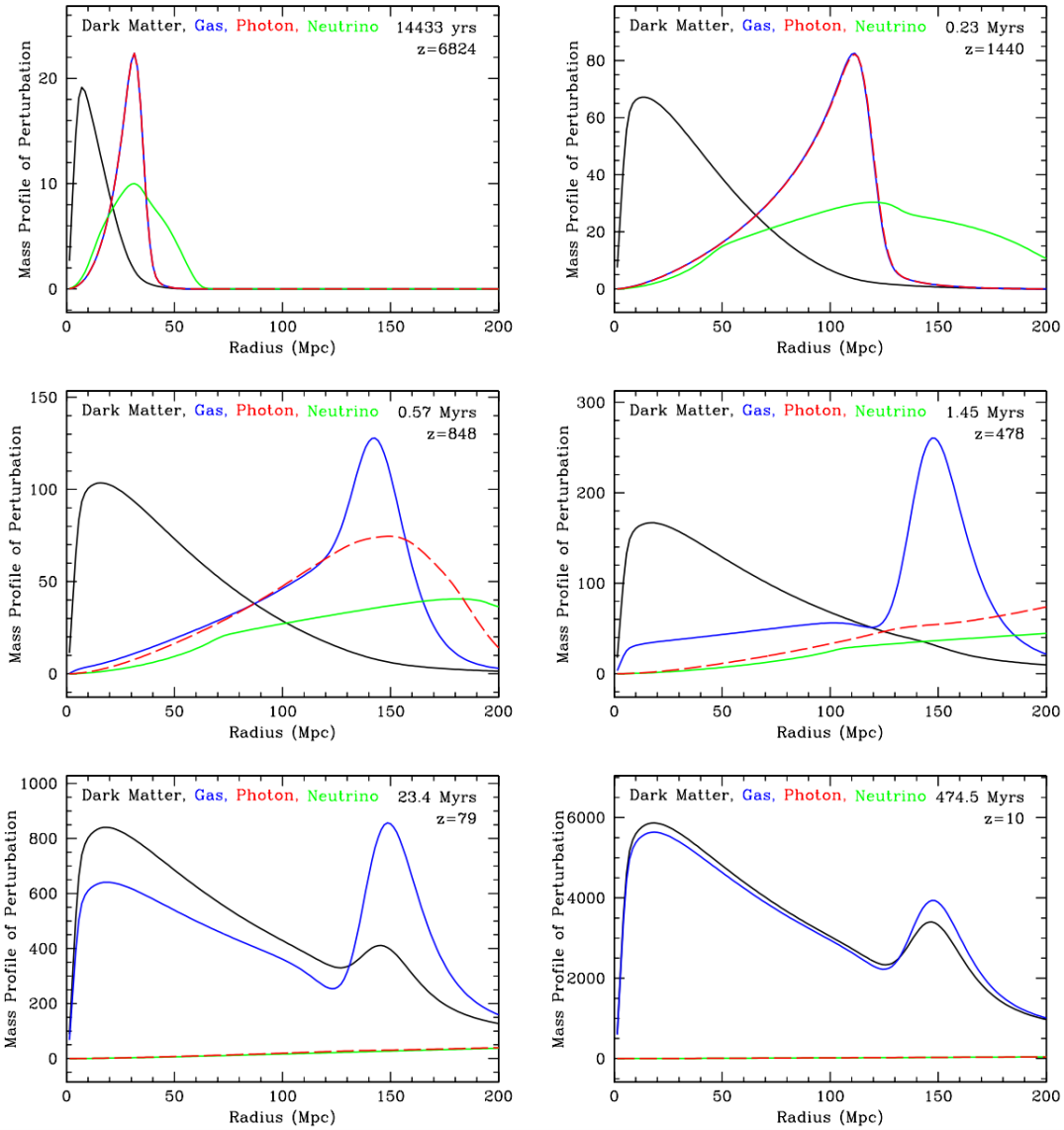


Figure 1.3: Illustration of the evolution of a spherical over-density from Eisenstein et al. (2007). Each figure shows increasing time/decreasing redshift. The curves within each figure correspond to a different component (dark matter, gas, photons and neutrinos). The x-axis is radial distance from the original over density, the y-axis is density. The top two plots show the wave of the photon-gas plasma expanding from the over-density. The central two plots show the photons uncoupling from the gas and continuing to expand. The final two plots show gravity realigning the dark matter and gas distributions.

instead corresponds to a sinusoid of a given frequency (as the Fourier transform of a peak is a sinusoid). The scale of the BAO is fixed once it is frozen in - the angular scale of this peak can be tracked over different redshifts to track the expansion rate over cosmic time. See Peebles & Yu (1970); Sunyaev & Zeldovich (1970); Bond & Efstathiou (1984) for early

discussions of the BAO, and Bassett & Hlozek (2010) for a modern review.

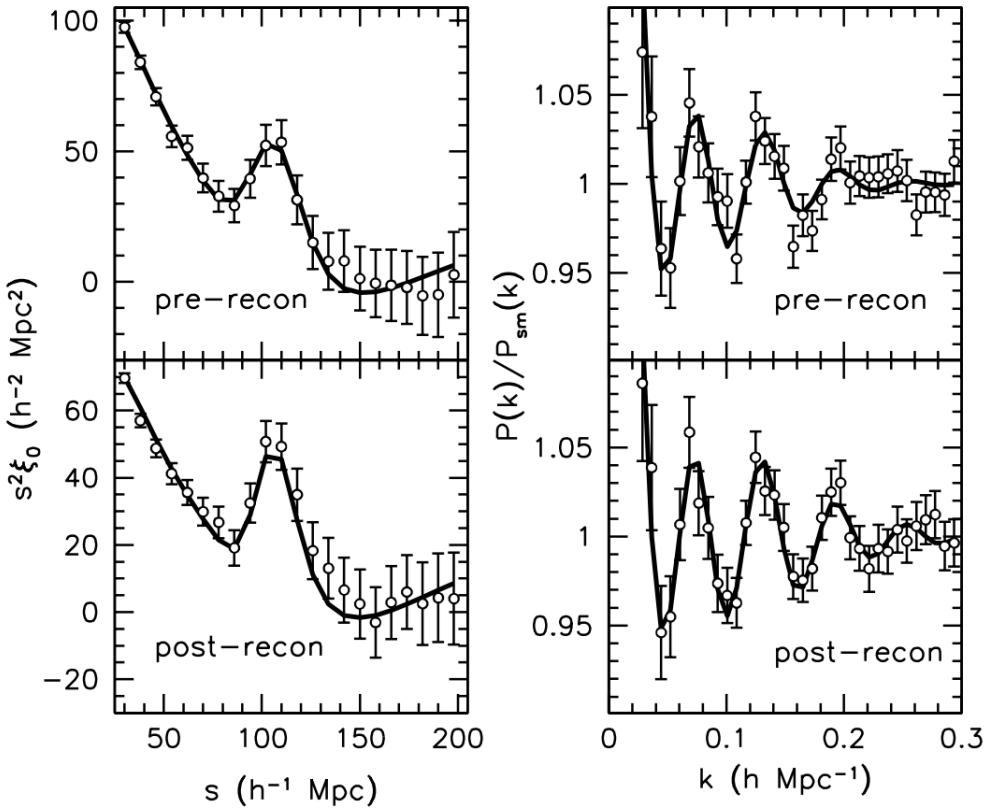


Figure 1.4: Measurements of the BAO in the SDSS BOSS survey from Anderson et al. (2014). The left two plots show the peak in the two point correlation function, and the right two plots show the corresponding sinusoid in the power spectrum/Fourier space (these plots are different representations of the same data). The top two plots show the BAO measured 'as is', the bottom two show it after the 'reconstruction' process, which corrects for small non-linearities, for more accurate measurements.

The actual value of the length scale of the BAO is the distance travelled by the acoustic waves over the time range spanning from the beginning of the Universe to time of photon-baryon decoupling:

$$s = \int_{z_{rec}}^{\infty} \frac{c_s \mathrm{d}z}{H(z)} = \frac{1}{\sqrt{\Omega_{m,0} H_0^2}} \frac{2c}{\sqrt{3 z_{rec} R_{rec}}} \log \left[\frac{\sqrt{1+R_{rec}} + \sqrt{R_{rec}+R_{eq}}}{1+\sqrt{R_{eq}}} \right]. \tag{1.14}$$

This works out as $\sim$ 150 Mpc, making it an extremely large standard ruler. The evolution of the angular distance as a function of redshift through BAO has now been

tracked over a range of redshifts, and is well understood both in real and Fourier space, and also in transverse and perpendicular scales (Percival et al., 2007, 2010).

1.1.4.4 Weak Lensing of Large-Scale Structure

The large-scale structure (LSS) of the Universe, the large scale distribution of matter, is a cosmological probe because it evolves differently in different cosmological models. Both BAO and the CMB are LSS-type measurements, but there are several other ways LSS can give cosmological information. In particular, the matter power spectrum (see section 1.2.1) evolves differently in different cosmological models, so obtaining direct access to the matter power spectrum gives cosmological information. This is in general hard to do because the matter power spectrum is dominated by dark matter, which doesn't emit light. It *is* possible to measure the galaxy power spectrum - but galaxies are biased tracers of matter, which complicates matters, one of the main themes of this thesis (see Chapter 3). Measuring the matter power spectrum by measuring the galaxy power spectrum requires untangling galaxy bias, see section 1.2.2.3.

One way in which we can measure the matter power spectrum is by measuring the deflection of light through gravitational lensing (for which the form of matter, dark or baryonic, doesn't matter). Light from background galaxies is lensed by a foreground matter distribution, imparting a distorting shear on the measured galaxy ellipticities, which can be statistically extracted to give constraints on the matter power spectrum. This effect is normally known as *weak lensing*[7] (see Bacon et al., 2000, and Kilbinger, 2015 for a review). Figure 1.5 illustrates how the shapes of background galaxies are distorted by weak lensing (and the corresponding inferred matter distribution).

There are many more LSS probes of cosmology that are beyond the scope of this introductory chapter e.g. the bispectrum (see Sefusatti et al., 2006), or massive galaxy cluster counts (large scale structure theory predicts that the number of massive galaxy clusters should be a function of cosmology[8], so counting clusters can give cosmological constraints,

[7] A related but distinct measurement is *galaxy-galaxy lensing*, which measures the galaxy-matter cross-power spectrum, as opposed to weak lensing, which is direct matter-matter power spectrum - although they both involve gravitational lensing in the weak field regime.

[8] See section 1.2.2.2, essentially the high mass end of the HMF.

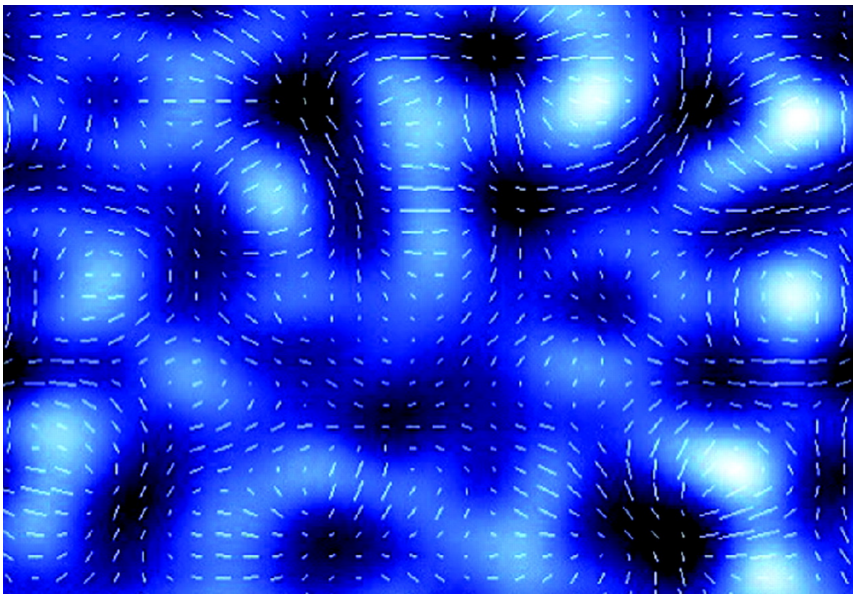


Figure 1.5: Illustration of weak gravitational lensing from Ellis (2010). White markers show the expected shear imparted on background galaxies (e.g. a source would be stretched/elongated along the length of the line). The blue shading represents the foreground matter distribution, from densest (white), to voids (black).

see Campanelli et al., 2012).

1.1.4.5 Joint Constraints

Cosmological models are typically described by several parameters; the measurements described thus far in this section can then be used to constrain these parameters. Typically however an individual probe will not be able to constrain all the parameters simultaneously e.g. in general some of the parameters will be degenerate with each other. This makes it necessary to use several of the aforementioned cosmological probes *simultaneously* to single out one isolated part of parameter space that is consistent with all observed data. Figure 1.6 shows how SNe, the CMB and the BAO are able to jointly constrain where in the $\Omega_M - \Omega_\Lambda$ plane our Universe lies. In particular note that even when the probe is 'precise' (like the CMB), the degeneracy between the parameters means that there would still be very large uncertainty if only that one experiment was used.

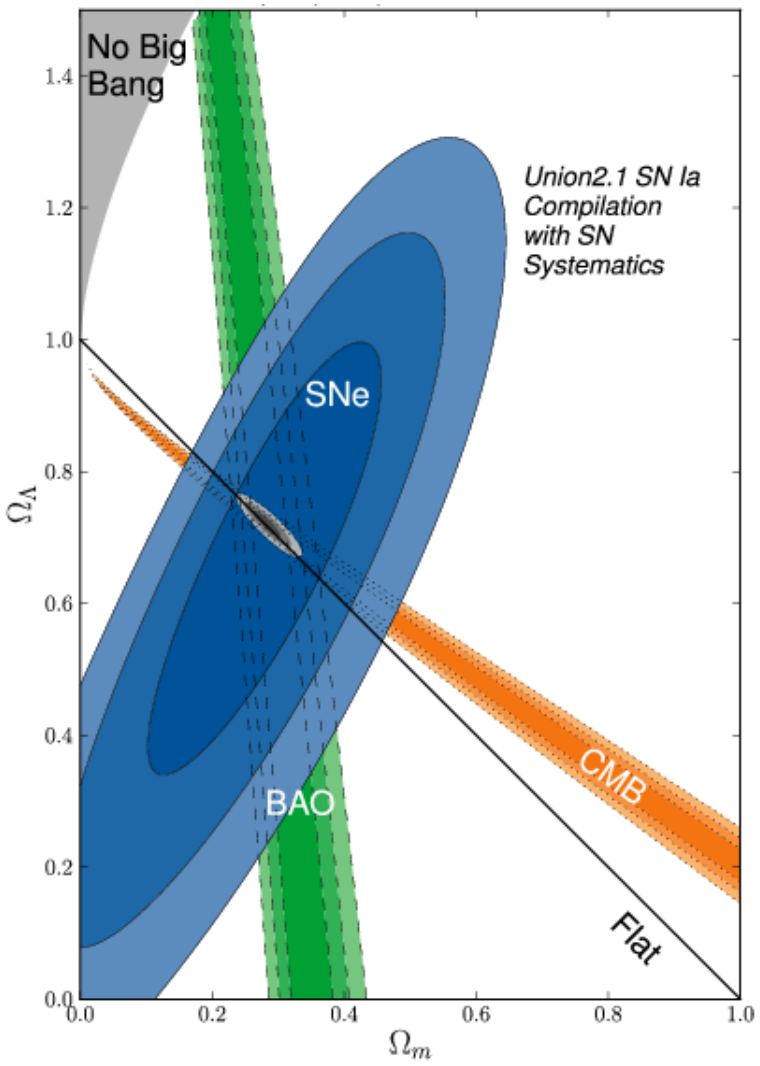


Figure 1.6: Figure from Suzuki et al. (2012) showing constraints on $\Omega_{M,0}$ (x-axis), and $\Omega_{\Lambda,0}$ (y-axis). Blue shading shows constraints from supernovae, green shows constraints from the BAO, orange shows constraints from the CMB, and the central small white/grey the constraints from the three simultaneously. The black line marks the space corresponding to a flat Universe (above the line corresponds to positive curvature, below negative curvature). The top left shaded area corresponds to a Universe that has always existed, with no Big Bang. It is evident that having multiple probes greatly helps in 'honing' in on one part of parameter space, corresponding to our familiar model of ΛCDM, $\Omega_{M,0} \approx 0.3$, $\Omega_{\Lambda,0} \approx 0.7$.

1.1.5 The Nature of Dark Matter

Cosmological observations, mainly measurements of large scale structure (see section 1.2), suggest that $\sim 80\%$ of the matter content of the Universe is composed of a pressure-free non-interacting form of matter - if all this mass was strongly interacting baryonic matter then structure would grow differently. However there are several other measurements that suggest that there really is a form of invisible matter permeating the Universe, which we briefly discuss here.

1.1.5.1 Evidence for Dark Matter

For many decades it has been known that for most galaxies, the total mass inferred from visible matter (e.g the sum of the mass in the visible stars and gas[9]) is substantially less than from estimates that directly measure the total mass. The most famous evidence for this discrepancy is galaxy rotation curves. The rotational velocity of stars (and gas as inferred from HI measurements) as a function of radius from a galaxy centre is much higher to large radii than would be expected from the gravitational forces from the visible matter alone - embedding these galaxies in vast dark matter globules larger than the individual galaxies solves this problem (see Rubin et al., 1980; Sofue & Rubin, 2001). Similarly, the total mass around galaxies inferred from gravitational lensing is far greater than the mass of the visible baryonic matter (see Wu et al., 1998; Natarajan et al., 2017) - and is consistent with the mass inferred from dynamics/rotational curves. This inconsistency between total mass, and baryonic mass, can be observed on scales larger than individual galaxies e.g. the velocity dispersion of galaxies in clusters implies a gravitational potential greater than the sum of the masses of the galaxies and gas in the cluster, see Zwicky (1937); Sohn et al. (2016). Thus there appears to be substantial evidence that galaxies are embedded in 'blobs' of dark matter, which is consistent with the predictions of large-scale structure (see section 1.2).

The 'bullet-cluster' (Clowe et al., 2004) is typically given as the 'smoking-gun' that dark matter really is the best explanation for observations (as opposed to alternative possibilities such as modifying gravity on galaxy scales etc., which is proposed by supporters of MOdified

[9]We believe we now have sufficient multi-wavelength observations and understanding of stellar processes to be confident that our estimates of the baryonic mass are reasonably accurate.

Figure 1.7: Composite image of the bullet cluster. The orange/white light is optical imaging from the Hubble Space Telescope, showing the galaxies, both in the cluster itself, and in the background. The red colouring is X-ray intensity from the Chandra space telescope, mapping hot gas. The blue shading is gravitational lensing potential, tracing where most of the mass is. The fact that the visible mass and total mass are misaligned is taken as a signal of the existence of dark matter. Image credit: NASA Astronomy Picture of the Day 24th August 2006.

Newtonian Dynamics, MOND, see Milgrom, 1983). See Figure 1.7 for a composite image of the Bullet-Cluster. When imaged at X-ray wavelengths, the cluster is observed to have a huge amount of X-ray emission in the centre, coming from hot gas. However most of the mass, as inferred from gravitational lensing, is on either side of the X-ray emission. Within the dark matter paradigm this has a straightforward interpretation. Two blobs of dark matter, each containing hot gas, collided. The dark matter simply passed through each other and came out the other side, as gravity was the only force acting on it. But the gas interacted and remained at the site of the collision, giving off X-ray emission. Modified gravity explanations typically struggle to explain these observations.

1.1.5.2 Different Dark Matter Candidates

Dark matter candidates can be broken down into 'hot', 'warm', and 'cold' dark matter (see Davis et al., 1985). These titles do not refer to a temperature per say - they correspond to the mass of the dark matter particle, which in turn determines the typical speed of the particles, and when they stopped behaving relativistically in the early Universe. Hot dark

matter describes candidates that are still relativistic in the Universe today e.g. neutrinos, see Primack (2001). Warm dark matter corresponds to particles of masses ∼keV that stopped behaving relativistically during the radiation dominated period of the Universe, with free-streaming lengths around the size of a proto-galaxy (Blumenthal et al., 1982; King & Merle, 2012). Cold dark matter describes candidates that have masses GeV or greater, that stopped being relativistic extremely early in the history of the Universe, and had free streaming scales much smaller than galaxy scales (Blumenthal et al., 1984). The majority of evidence suggests that dark matter has to be cold (Frenk & White, 2012), although some authors suggest warm-cold hybrid models are just about still permitted by the data (e.g. Kamada et al., 2016). If dark matter was hot, then fluctuations on small scales are quickly wiped out. However if cold, then dark matter can form fluctuations on arbitrarily small scales. See Figure 1.8 for an illustration of the differences for how structure forms depends on the type of dark matter present in the Universe (section 1.2 discusses in greater detail how we believe structure grows). Figure 1.8 shows visualisations of the structure of dark matter in three simulations with the three different forms of dark matter. The simulation of cold dark matter shows very granular structure, with the 'cosmic web' visible down to very small scales. In contrast the hot dark matter simulation is very smooth on very small scales - one has to go to very large scales before anisotropy is visible.

Cold dark matter candidates are then further divided into 'Weakly Interacting Massive Particles' (WIMPs), or 'MAssive Compact Halo Objects' (MACHOs). WIMP candidates are new particles beyond the standard model of particle physics. Typical MACHO candidate models propose that galaxies have large numbers of unobserved planet to star massed objects e.g. the unobserved mass is comprised on large numbers of orphaned planets, or black holes that aren't accreting any matter (and hence no light). Micro-lensing experiments have largely ruled out MACHOs of masses in the range $10^{-7}M_{\odot} < M < 15M_{\odot}$ (Tisserand et al., 2007). Furthermore, independently, many extensions to the standard model of particle physics seem to require the existence of new particles with close to the right properties. The current consensus is that dark matter is a WIMP - astronomy eagerly awaits either a) the production of a new particle with the appropriate properties in a particle accelerator like the Large Hadron Collider (Mitsou, 2015), b) detection of dark matter in neutrino telescopes

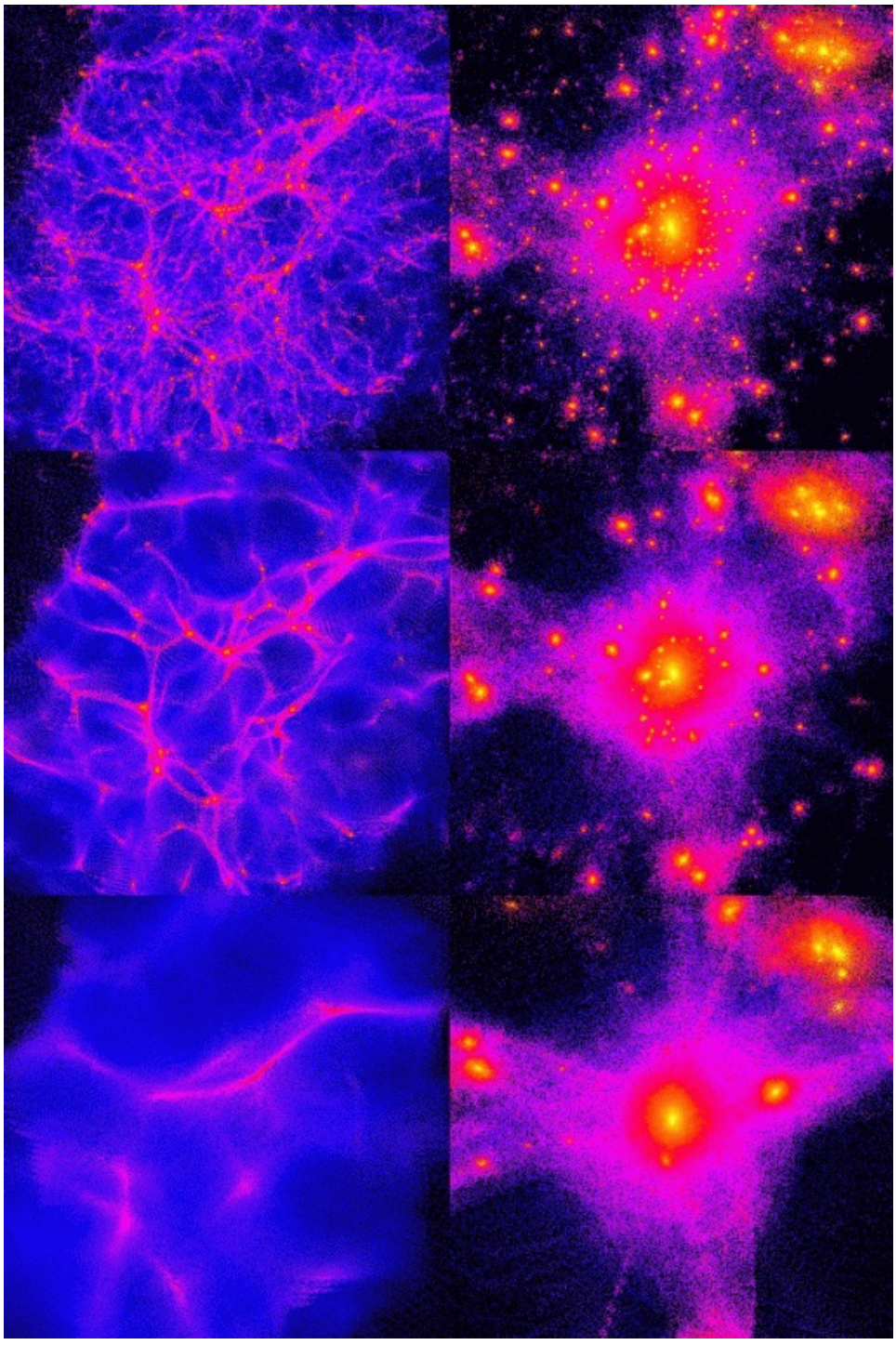


Figure 1.8: Simulated illustration of how structure would form differently if dark matter was cold (top row), warm (middle) and hot (bottom) respectively. Pink corresponds to dense regions, blue to under-dense regions. The right hand figures show 'zoomed in' regions of the left hand figures. Image credit, ITP, University of Zurich.

like Super Kamiokande (Kappl & Winkler, 2011) or c) evidence of dark matter annihilation in massive clusters in gamma ray observations[10] (e.g. Mambrini & Toma, 2015).

1.2 Structure Formation

The FLRW metric described in section 1.1.1 assumes a homogeneous and isotropic Universe. However we believe our Universe is only homogeneous on *large scales* - and as discussed in section 1.1.4.4, the growth of inhomogeneities is one of our main ways of learning about cosmology. In this section we discuss how inhomogeneities form and grow over the history of the Universe.

1.2.1 Linear Perturbation Theory

For small inhomogeneities on large scales it is possible to understand the growth of structure analytically. Linear Perturbation Theory accomplishes this by considering small deviations from FLRW:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}, \tag{1.15}$$

where $g_{\mu\nu}$ is the metric of the perturbed space time, $\bar{g}_{\mu\nu}$ is the FLRW metric, and $\delta g_{\mu\nu}$ is a small perturbation to FLRW. There are many different ways of expressing the same physical perturbations, which are called gauges. It is possible to fix some of these degrees of freedom; in the absence of anisotropic stress, and for perturbations that decay at large scales, one can uniquely go to the Newtonian Gauge:

$$\mathrm{d}s^2 = a^2(\tau)\left[(1+2\psi)\mathrm{d}\tau^2 - (1-2\psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right], \tag{1.16}$$

where τ is conformal time (a reparametrization of the usual time coordinate, $\mathrm{d}\tau = \frac{\mathrm{d}t}{a(t)}$), and ψ is a function of space and time and is small. This effectively describes the perturbations in such a way that dense regions have experienced less time, and under-dense

[10] Most potential candidate particles from particle physics don't have an interaction rate of zero, just extremely low rates, so it is possible that in the densest dark matter regions it could be feasible to detect the particles annihilating.

regions have experienced more time.

The theory of linear perturbation evolution is well developed. A core property of the theory is that the fluctuations on different scales, when expressed in Fourier space, grow *independently* of each other i.e. different modes don't couple with each other. This means that the evolution of each mode can be calculated independently, normally satisfying some differential equation. In a ΛCDM Universe, the evolution of linear perturbations of matter are well understood, and behave qualitatively different in the limits of i) the three separate stages of cosmic expansion, radiation, matter and dark energy dominated and ii) super-horizon and sub-horizon spatial scales. Super/sub-horizon refers to scales that are larger/smaller than the Hubble horizon ($1/H$), the scale beyond which particles moving with the Hubble flow appear to be moving faster/slower than the speed of light.

We do not cover the details here, but it can be shown that when the Universe is radiation dominated, on super-horizon scales $\delta_{\rm m}(k) \propto a^2$, and on sub-horizon scales $\delta_{\rm m}(k) \propto \log a$. So sub-horizon scales can only grow extremely slowly when the Universe is radiation dominated. When the Universe is matter dominated, perturbations grow $\delta_{\rm m}(k) \propto a$ on all scales, and when the Universe is dark energy dominated, perturbations do not grow. The growth that a mode receives is often parametrised with a 'transfer function' $T(k)$. During radiation and matter dominated phases of the Universe, the Hubble Horizon is always increasing. Before matter-radiation equality, scales larger than the horizon are all growing at the same rate, but scales smaller than the horizon are essentially not growing at all. Modes continue to grow until the horizon reaches their scale, and then they are 'frozen' in, and stop growing. Then eventually matter-radiation equality is reached, and all scales grow at the same rate. Thus for scales larger than the Hubble horizon at matter-radiation equality ($k_{\rm eq} \sim 0.01{\rm Mpc}^{-1}$), all modes have grown at the same rate for all time and $T(k) \propto 1$ for $k < k_{\rm eq}$. For scales smaller, the modes must have entered the horizon at some point before matter-radiation equality - with smaller scales entering the horizon first. Consider two modes $k_1, k_2 > k_{\rm eq}$, with wavelengths λ_1, λ_2, that enter the horizon at scale factors a_1, a_2, and have transfer functions T_1, T_2. We straightforwardly have $a_1/a_2 = \lambda_1/\lambda_2$, and $a_1/a_2 = k_2/k_1$. Until the mode enters the horizon it grows $\propto a^2$, so $T_1/T_2 = (a_1/a_2)^2$. Combining these, we can see that $T_1/T_2 = (k_2/k_1)^2$, and thus $T(k) \propto 1/k^2$.

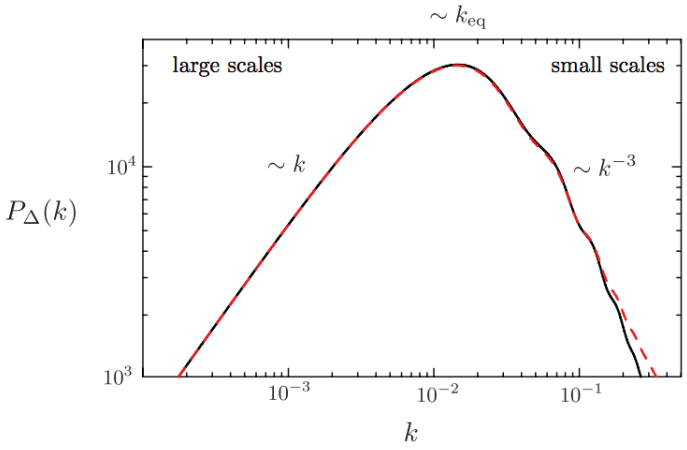


Figure 1.9: The model matter power spectrum from linear theory (black line), with non-linear corrections (red dashed line, see section 1.2.2). Note the rough power law behaviour either side of $k_{\rm eq}$ The oscillations on small scales are baryonic acoustic oscillations, section 1.1.4.3. Figure from Daniel Baumann Part III Cambridge Cosmology notes.

This gives us:

$$T(k) \propto \begin{cases} 1, & k < k_{\rm eq} \\ k^{-2}, & k > k_{\rm eq} \end{cases} \tag{1.17}$$

Our measurable quantity of physical interest is the power spectrum, $P(k,z) \propto \delta(k,z)^2$. Therefore the matter power spectrum today, $P(k, z = 0)$, can be expressed as $P(k) = T^2(k, z = 0)P_{\rm init}(k)$, where $P_{\rm init}(k)$ is the initial power spectrum. It is postulated that the power spectrum was initially in the form of a power law $P_{\rm init}(k) \propto k^n$, with $n \approx 1$ (see Liddle & Lyth, 2000). This means that the matter power spectrum today must be:

$$P(k) \propto \begin{cases} k, & k < k_{\rm eq} \\ k^{-3}, & k > k_{\rm eq} \end{cases} \tag{1.18}$$

This can be calculated to greater precision numerically, see figure 1.9.

Note that in a non-expanding Universe, linear perturbations would grow exponentially, so the fact that they only grow as power laws can be interpreted as the expansion of the Universe 'fighting' gravity making the modes grow. When the Universe starts accelerating exponentially, modes do not grow at all, which can be interpreted as Hubble flow taking matter away from each other faster than gravity can 'travel', meaning modes cannot grow.

1.2.2 The Halo Model

The theory described in section 1.2.1 describes linear perturbations to the cosmic density field. However this theory is only valid for small perturbations, and typically fails once perturbations grow too large for the linear approximation to be accurate. Both analytic models and simulations suggest that once over-densities become too large they collapse into discrete over-densities of dark matter known as haloes. These haloes are believed to be crucial to galaxy formation and are a subject of great interest in large-scale structure.

We discuss several toy/analytic models here that describe a) how haloes form (spherical collapse model, section 1.2.2.1), b) how many of these haloes form (Press-Schechter Theory, section 1.2.2.2), and c) where they form relative to large-scale linear over densities (Peak Background Split, section 1.2.2.3). These models are believed to *qualitatively* capture features about the halo population, but the behaviour is inherently non-linear and hard to characterise analytically - it is necessary to do N-body simulations (section 1.2.2.4) to obtain precise quantitive information about the halo population.

1.2.2.1 Spherical Collapse Model

The spherical collapse model is an analytic 'toy model' that can give insight to how small linear over-densities evolve into non-linear haloes (Gunn & Gott, 1972; Peebles, 1980; Lahav et al., 1991). If we consider a sphere of radius r, of slightly raised density over an otherwise uniform density (matter dominated) Universe, then Gauss's law says that the only relevant mass for gravitational purposes for particles at the edge of the sphere is the mass inside. Thus we can write:

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{-GM}{r^2}, \tag{1.19}$$

where M is the total mass of the over-density, r is the radius of the over-density, and t is time. This integrates to

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{2GM}{r} + C, \tag{1.20}$$

which is effectively conservation of energy. This has the parametric solution[11]:

$$r = A(1 - \cos\theta), \tag{1.21}$$

$$t = B(\theta - \sin\theta), \tag{1.22}$$

where $A^3 = GMB^2$, $C = -A^2/B^2$, and θ is a reparametrisation of time. We have $C < 0$ for collapse (potential energy greater than kinetic energy). For small t (or θ), this can be solved to find $r \propto t^{2/3}$, or in other words the overdensity follows the Hubble flow and expands with it.

We can then do some algebraic manipulation to find the next order expansion:

$$r = \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right], \tag{1.23}$$

e.g. the over-density expands with the Hubble flow up to $\theta = \pi$, before collapsing back towards zero at $\theta = 2\pi$. We can connect this to the over-density δ by considering $M = (4\pi/3)\bar{\rho}(1 + \delta)$. First order expansion then gives

$$\delta \approx \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{1.24}$$

Substituting in the values of theta for peak radius ('turnaround'), and for when the over-density has collapsed back to zero radius, it can be seen that

$$\delta_{turnaround} \approx \frac{3}{20}(6\pi)^{2/3} = 1.06, \tag{1.25}$$

$$\delta_{collapse} \approx \frac{3}{20}(12\pi)^{2/3} = 1.69, \tag{1.26}$$

so over-densities turnaround and start collapsing at $\delta \approx 1$, and end with a collapse over-density of $\delta \approx 1.7$. This $\delta_{collapse}$ is slightly artificial as in a completely idealised spherical overdensity it will collapse to ∞, the $\delta \approx 1.7$ is the value in the linear approximation.

[11] which is identical to the solution of the Friedmann equations for a closed matter dominated Universe

Nonetheless it turns out that this is a useful figure for identifying when over-densities collapse.

In practice the over-density does not collapse to zero; small asymmetries cause it to virialise into a roughly spherically symmetric 'halo' before then. If potential energy is U, kinetic energy is K, and total energy is E, then we have $E = U + K$, and the virial theorem says $U = -2K$. This implies $E = -K = U/2$ for the final state. We have $U \propto r^{-1}$, and potential energy is at its maximum at turnaround. So $r_{vir} = r_{turnaround}/2$, at $\theta = 3\pi/2$. If we want to understand the over-density of virialised objects we must compare mean density at collapse, and the final density of the over-density at virialisation. This can be shown to correspond to $r = A$ at virialisation, with density $\rho_h = 3M/4\pi A^3$. In a matter dominated Universe, mean density goes as $\bar{\rho} = (6\pi G t^2)^{-1}$, and time at collapse is $t = 2\pi B$, giving:

$$\frac{\rho(\theta = 3\pi/2)}{\bar{\rho}(\theta = 2\pi)} = 18\pi^2 \approx 178 \tag{1.27}$$

Thus all collapsed objects have over-densities of about $\Delta \approx 178$, which is a cosmology dependent figure. Typically simulations for ΛCDM find that $\Delta \approx 200$ is most appropriate for identifying virialised overdensities.

1.2.2.2 Press-Schechter Theory - Halo Mass Function

A natural question to ask about haloes is how many of them form, and what distribution of masses do they have. This is quantified through the comoving number density of haloes, per unit of halo mass - known as the *halo mass function*. To calculate this, we can consider overdensities that surpass $\delta_c = 1.69$. Analytic halo mass function calculations were first laid out by Press & Schechter (1974), whose approach we briefly summarise.

Firstly for masses M, we can smooth the density field on the associated scale R, from $M = (4/3)\pi R^3$. This results in a Gaussian field of variance $\sigma(M)$.

Then the fraction of matter in the Universe that is collapsed can be found by integrating a Gaussian distribution:

$$f(> M) = \frac{1}{\sqrt{2\pi}} \int_{\delta_c/\sigma(M)}^{\infty} \mathrm{d}x \, e^{-x^2/2}. \tag{1.28}$$

It is convenient to define $\nu = \delta_c/\sigma(M)$ as an alternative paramatisation of halo mass. Then, to find the fraction of mass between M and $M + dM$, we take the derivative:

$$\frac{df}{dM} = \frac{1}{\sqrt{2\pi}}\left[\frac{dx}{dM}e^{-x^2/2}\right]_{x=\nu} = \frac{1}{\sqrt{2\pi}}\frac{d\nu}{dM}e^{-\nu^2/2}. \tag{1.29}$$

To convert to number density between M and $M + dM$ we need to multiply by ρ_0/M:

$$\frac{dn}{dM} = \frac{\rho_0}{M}\frac{df}{dM} = \frac{\rho_0}{M}\frac{1}{\sqrt{2\pi}}\frac{d\nu}{dM}e^{-\nu^2/2}. \tag{1.30}$$

Then making the substitution $\frac{d\log\nu}{d\log M} = \frac{M}{\nu}\frac{d\nu}{dM}$ we get:

$$\frac{1}{M}\frac{dn}{dM} = \frac{dn}{d\log M} = \frac{\rho_0}{M}\frac{1}{\sqrt{2\pi}}\nu e^{-\nu^2/2}\frac{d\log\nu}{d\log M}. \tag{1.31}$$

To evaluate $\frac{d\log\nu}{d\log M}$, we need to consider $\sigma(M(R)) = \sigma(R)$. This can be found by relating the power spectrum and the spherical top hat filter:

$$\sigma(R) = \int_0^\infty \frac{k^2dk}{2\pi^2}P(k)W(k,r)^2 = \int_0^\infty \frac{k^(2+n)dk}{2\pi^2}\left(\frac{3\sin(kr) - 3kr\cos(kr)}{(kr)^3}\right)^2, \tag{1.32}$$

where $W(k,r) = \frac{3\sin(kr)-3kr\cos(kr)}{(kr)^3}$ and $P(k) \propto k^n$. In the limiting cases, we have $W(k,r) \approx 1$ for $kr \to 0$ and $W(k,r) \approx 0$ for $kr \to \infty$, so we can approximate the term inside the brackets as a step function equal to 1 up to $k = 1/r$, and equal to zero for higher values of k. This approximation yields:

$$\sigma(R) = \frac{1}{2\pi^2}\frac{1}{r^{n+3}}. \tag{1.33}$$

So $\sigma \propto r^{-n+3} \propto M^{-(n+3)/6}$ and $\nu = \delta_c/\sigma(R) \propto M^{(n+3)/6}$. This means $\frac{d\log\nu}{d\log M} = (n+3)/6$ (where n is the power spectrum index at the scale k that corresponds to mass M). Also at this point Press & Schechter (1974) introduce an extra 'fudge-factor' of 2. Extrapolating the earlier $f(> M)$ down to low masses would suggest that only half of matter is in haloes, which is not correct[12]. This extra factor gives:

[12]Essentially what is happening is that this toy model only includes the collapse of over-densities, it doesn't include the formation of under-densities, so an extra factor of 2 is needed.

$$\frac{dn}{d\log M} = \frac{\rho_0}{M}\sqrt{\frac{2}{\pi}}\left(\frac{n+3}{6}\right)\nu e^{-\nu^2/2}. \tag{1.34}$$

If we introduce a characteristic halo mass $M_\star$ defined by $\nu = (M/M^\star)^{(n+3)/6}$ we finally get:

$$\frac{dn}{dM} = \frac{\rho_0}{M^2}\sqrt{\frac{2}{\pi}}(M/M^\star)^{(n+3)/6}\exp\left[-\left(\frac{M}{M^\star}\right)^{(n+3)/6}\right]. \tag{1.35}$$

This is the halo mass function. For high mass we go to small k values, $n \to 1$, and the mass function cuts off exponentially. For low masses, we go to large k values, $n \to -3$, so the mass function goes as M^{-2}. This completes our qualitative view of the halo mass function - a power law at low halo masses, with an exponential cut off at some critical mass $M^\star$. $M^\star \approx 10^{13} M_\odot$ in the Universe today.

One way to describe Press-Schechter theory is that it essentially assumes linear theory is perfectly correct up to $\delta = 1.69$, and then haloes instantaneously collapse to $\delta = 200$ straightaway and remain there.

Since the original formulation in Press & Schechter (1974) various more sophisticated formulations have been devised, e.g. Sheth et al. (2001) - but the basic qualitative idea has not changed. The most precise halo mass functions are now considered to come from N-body simulations, see section 1.2.2.4.

1.2.2.3 Peak Background Split - Halo Bias

If haloes formed independently of the local density fluctuation, then haloes would perfectly trace the large scale dark matter distribution. However this is not the case - as we will see in this section, haloes preferentially form in denser regions of the Universe. To quantify this it is often useful to consider a halo 'bias':

$$\left(\frac{\delta n}{n}\right) = b\left(\frac{\delta\rho}{\rho}\right), \tag{1.36}$$

where n is the number of haloes and ρ is the dark matter density, and b is the bias. For example, a bias of 2 would mean that a 10% matter over density would correspond to a

20% halo over-density. The bias of a galaxy population is defined in an analogous manner. Understanding bias is a major topic of this thesis. Essentially large scale density can act as an enhancement in dense regions, making it 'easier' to reach the threshold of over-density to form a halo. The formalism we use follows the approach described in Cole & Kaiser (1989); Efstathiou et al. (1988) and Mo et al. (1997), which originally laid out the theory.

The first step is to consider over-density 'split' into a local over-density and a larger-scale over-density:

$$\delta = \delta_b + \delta_p, \tag{1.37}$$

where δ_b is the background over-density, and δ_p is the peak over-density. This effectively *lowers* how high an over-density has to be to collapse: $\delta_{collapse} = \delta_{critical} - \delta_b$, e.g. the over-density required for a collapse is reduced, permitting more haloes to form locally.

If we Taylor expand the number density in δ_b:

$$n_M(\delta_b) = n_M^{mean} + \frac{dn_M}{d\nu}\frac{d\nu}{d\delta_b}\delta_b..., \tag{1.38}$$

where n_M^{mean} is the global mean number density, and $n_M(\delta_b)$ is the local number density. We use the Press-Schechter halo mass function we have $n_M \propto \nu \exp(-\nu^2/2)$, corresponding to $\frac{dn_M}{d\nu} = \frac{\nu^2-1}{\nu} n_M$, and $d\nu/d\delta_b = 1/\sigma$, giving us:

$$n_M(\delta_b) \approx n_M \left(1 + \frac{\nu^2 - 1}{\sigma\nu}\delta_b\right), \tag{1.39}$$

which corresponds to

$$b(M) = 1 + \frac{\nu^2 - 1}{\delta_c}. \tag{1.40}$$

This gives a qualitative view of bias being near 1 at low to intermediate halo masses, and then growing increasingly rapidly with halo mass at the high mass end. Extremely low mass haloes are 'anti-biased' ($b < 1$) e.g. less clustered than dark matter as a whole. 'Voids' (areas of extreme low density) are also measured to be 'anti-biased'. Furthermore, although the above expression has no redshift dependence, ν as a function of halo mass

changes with redshift. In particular, a fixed halo mass corresponds to lower values of ν as redshift decreases, so the bias of haloes of a given mass decreases over time.

Again, as with the Press-Schechter mass function, there has been extensive work to improve our analytic understanding of halo bias (e.g. Tinker et al., 2010), but the most precise estimates of bias come from N-body simulations.

1.2.2.4 N-Body Simulations

The evolution of dark matter in a FLRW Universe is in principle a well understood process - without pressure, collisions or baryonic physics, the only relevant force is gravity. In addition we believe we understand the initial conditions well, which makes a dark matter only Universe easy to simulate - typically known as *N-body simulations*, where the positions of 'N' particles of a given mass are allowed to evolve under gravity. N-body cosmological simulations today typically have hundreds of billions of particles, probing tens of cubic Gigaparsecs, with a resolution sensitive to haloes of mass $10^{11} M_{\odot}$ e.g. Kim et al. (2011).

Cosmological N-body simulations are believed to give us the best insight into the large-scale distribution of dark matter. In particular, as already discussed, halo mass functions and halo biases from N-body simulations are believed to be more realistic. Typically after the simulation is run, some 'post-processing' is required for interpretation. The simulation is normally 'saved' at several different points in time, to give 'snapshots' throughout cosmic time. At each snapshot, a 'halo-finder' (see Knebe et al., 2011 for a comparison for different techniques to do this) is used to identify haloes. Then the fact that individual particles can be tracked between snapshots allows a halo merger tree to be constructed. Merger trees from N-body simulations are one of our few ways of obtaining 'temporal' information about dark matter - observations through gravitational lensing etc. can estimate the mass of a halo, but typically struggle to give information about the history of the halo.

Other key types of cosmological simulation are 'semi-analytic models' (SAMs), and 'hydrodynamical' simulations (which we discuss and compare to observations in Chapter 6). SAMs take a pre-existing N-body simulation, and overlay a galaxy population according to some analytic formulation, typically accounting for dark matter halo merger trees, total gas within each halo etc. Notable SAMs include Galform (Bower et al., 2006) and the model

of Guo et al. (2012). Hydrodynamic simulations seek to directly simulate the important physics for structure and galaxy formation, directly modelling gravity, hydrodynamics, and thermodynamics simultaneously. Hydrodynamic simulations typically model dark matter, gas and stars, normally alongside physical processes like star formation. Notable examples include Horizon-AGN (Dubois et al., 2014), the Evolution and Assembly of GaLaxies and their Environments project (EAGLE, Schaye et al., 2015; McAlpine et al., 2016) and Illustris (Vogelsberger et al., 2014).

1.2.2.5 Halo Profiles

One specific prediction from N-body simulations is that dark matter haloes typically obey a universal spherically symmetric profile first identified by Navarro et al. (1996), now known as the Navarro-Frenk-White (NFW) profile:

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2}, \tag{1.41}$$

where r is radius from the centre of the halo, ρ_0 is some normalising density, and R_s is some characteristic radius. The density of this profile behaves as $\rho \propto r^{-1}$ on small scales, and $\rho \propto r^{-3}$ on large scales, with the transition on the scale radius R_s. Often a second scale, the virial radius R_{vir}, is identified, the scale at which the density within this radius is 200 times the critical density (see section 1.1.1). Then one can define a 'compactness ratio' c, where $R_{vir} = c \times R_s$. The compactness ratio in general has a halo mass and redshift dependence, which can be measured in N-body simulations.

Some authors claim slightly different parameterisations provide a slightly superior fit to halo profiles in N-body simulations e.g. the Einasto profile (Einasto, 1965; Merritt et al., 2006), but typically the differences are very small - all modern simulations are in good agreement for halo profiles. Comparisons with observations presents slightly more tension. Profiles from N-body simulations are typically 'cuspy', in that the density asymptotes up at centre of the halo (reflected in the NFW profile density tending to infinity at $r = 0$), whereas observations (typically from gravitational lensing) ordinarily find profiles that agree with NFW on large scales, but have 'cores' at the centre, the density profile is flatter at the

centre of the halo. This is known as the 'cusp-core' problem (see de Blok, 2010). Although the problem is not completely resolved, it appears that baryonic physics can resolve some of the discrepancy, as galaxies become the dominant form of mass at the centre of haloes (Del Popolo & Pace, 2016). For the purposes of this thesis, we mainly consider NFW profiles, as the difference between profiles is negligible for our purposes. In addition, much of our modelling requires Fourier transforms for the profiles, and NFW profiles have analytic Fourier transforms.

1.3 Galaxy Evolution

The Milky Way, the galaxy we live in, appears as a fuzzy white 'band' over the sky to the naked eye on a clear night away from cities, and has been known since antiquity. Aristotle originally believed that the Milky Way was a terrestrial phenomenon associated with the Earth's atmosphere. By the 10th to 14th Centuries, Arabic and Persian astronomers were beginning to argue that the Milky Way was a myriad of distant stars. Finally the astronomical use of the telescope - first performed by Galileo - confirmed this hypothesis: "It is nothing else but a mass of innumerable stars planted together in clusters." he declared. The philosopher Immanuel Kant made several early contributions to the development of the concept of a galaxy, being the first to suggest that the Milky Way might be stars moving in orbits under gravity, and the first to suggest that observed cloud like nebulae might be galaxies, or 'island universes' as he described them, in their own right, as opposed to being part of the Milky Way. Confirming that this was correct had to wait until the 20th Century, where, as discussed, Hubble was able to show that nebulae were at vast distances and could not be part of the Milky Way. Since then astronomers have been able to observe huge samples of hundreds of thousands to millions of galaxies.

The astronomical study of galaxies is thus divided into a) 'Galactic studies' of the Milky way (and to some extent the nearest few galaxies to us, the Magellanic Clouds and the Andromeda Galaxy), which can be studied in exquisite detail due to the proximity of being inside it, and b) 'extragalactic astronomy', the study of other galaxies, where in general far less can be learned about each individual galaxy, but the full range of the galaxy population

can be studied as opposed to the sample size of one that the Milky Way gives. In this thesis I am predominantly concerned with extragalactic astronomy, and how galaxies formed and evolved over the history of the Universe.

1.3.1 Anatomy of a Galaxy Spectrum

Galaxies give out light over a range of wavelengths. In general accessing different parts of the electromagnetic spectrum gives access to information about different aspects of the galaxy - different populations of stars, different components (dust, gas, stars) etc. Having access to a range of wavelengths is also extremely important for high redshift sources e.g. if you want to observe the optical light from high redshift galaxies, you have to observe them at infrared wavelengths in the Earth's rest frame. The Earth's atmosphere is opaque for large sections of the electromagnetic spectrum, necessitating the use of space telescopes at these wavelengths - see Figure 1.10.

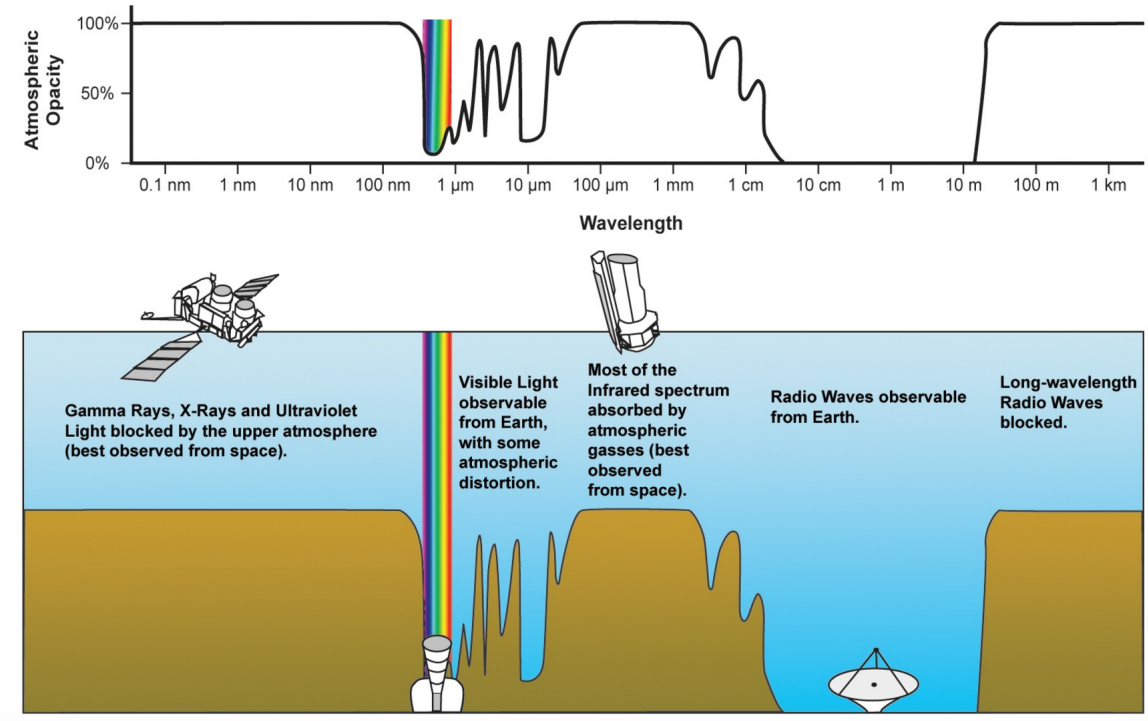


Figure 1.10: Cartoon from Humboldt State Geospatial Online showing which parts of the electromagnetic spectrum the atmosphere is transparent to.

1.3.1.1 Ultra-Violet, Optical and Infrared

Individual stars give off most of their light as black body thermal radiation in a relatively small range of wavelengths in the UV and optical. In general the trend is more (less) massive corresponds to hotter (colder) / bluer (redder) / brighter (fainter) / rarer (more common) / younger (older). Galaxies however have stars of a range of masses, so give out light at a broader range of wavelengths. Photons over 0.1-1.0 μm are mainly thermal light from stars. Many galaxies also have a substantial dust component - particles ranging in size from a single molecule up to a micrometre, typically formed from ejecta from supernovae (Draine, 2003). Dust absorbs UV photons strongly. Typically the UV photons heat up the dust, which then achieves a temperature of approximately 10-100 K, and in turn this dust emits an approximate black body spectrum at around 100 μm. The net effect of this 'reprocessing' is a reduction in the UV luminosity and a increase in the far infrared luminosity, as shown in figure 1.11.

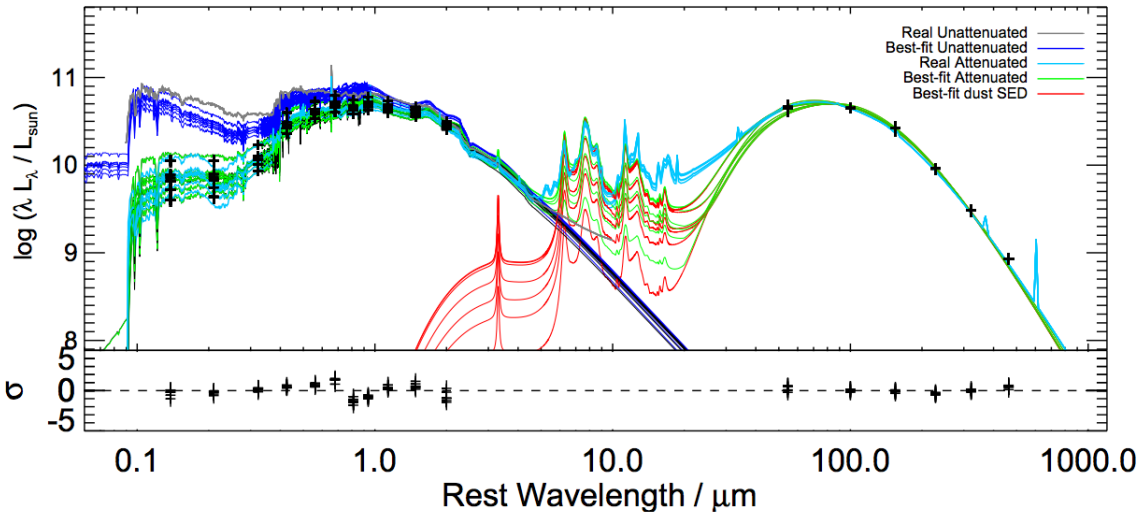


Figure 1.11: Example galaxy spectrum showing the two thermal peaks of stellar light and dust thermal emission. The black points are observations, the red curves are dust models, the blue curves are stellar thermal emission before dust absorption, the teal is stellar emission after dust, note the particular drop after absorption in the UV at $\sim 0.1\mu$m. Figure from Hayward & Smith (2014).

1.3.1.2 Radio

Radio emission (frequencies below a few GHz) from galaxies is dominated by synchrotron emission from active galactic nuclei (see section 1.3.6) and star formation. Electrons are accelerated by magnetic fields from the radio jets in the case of AGN, and by magnetic fields from supernovae remnants in the case of star formation, to give off synchrotron emission that follows a rough power law with frequency. Galaxies also often have some free-free emission, which has a flat spectrum. See Condon (1992) for a review.

1.3.1.3 X-ray

The fact that most galaxies are observed to have some X-ray emission only became apparent with the advent of space X-ray telescopes, as the Earth's atmosphere is highly opaque to X-rays, see Barcons (2001) for review of the use of X-ray sources in the study of cosmology and large-scale structure. Normal stars give off an extremely small fraction of the X-ray emission of galaxies. Compact binaries and supernovae remnants can give slightly larger contributions (Wang, 2011). The stellar component is normally sub-dominant to AGN (see 1.3.6) contributions and emission from the hot gas (Sarazin, 1986).

The hot gas in large haloes typically has a temperature of approximately $10^6 - 10^8$K, which corresponds to a thermal spectrum peaking at X-ray wavelengths. This emission is typically extended over the whole host halo of the galaxy, a much larger area than the galaxy.

1.3.1.4 Stellar Population Synthesis Models

In general, the observed spectrum of a galaxy is derived from the superposition of spectra of stars with a range of masses and ages. Stellar population synthesis (SPS) models seek to model this to unravel the star formation history from a galaxy spectrum - see Bruzual & Charlot (2003) for a more in depth description of one popular implementation of SPS modelling. We briefly summarise the process here. In general, a SPS model needs the following to be defined:

- A star formation history (e.g. total stellar mass formed per year as a function of time)

- An initial mass-function (distribution of stellar masses of stars when new stars formed)
- Star spectra as a function of age, metallically and mass
- A dust model

Using a combination of these a model galaxy spectrum can then be simulated. Much of the difficulty is finding the correct physics to input. Stellar physics is sufficiently well understood to generate good stellar spectra libraries (e.g. Le Borgne et al., 2003). There is still moderate uncertainty in the initial mass function, which cannot be predicted from theory, only from observations from Milky Way star bursts or stellar mass spectral indicators in other galaxies (see Bastian et al., 2010 for a discussion about current constraints on the initial mass function and its universality). Dust extinction also remains difficult to model (see Fischera et al., 2003). Nonetheless, SPS models have proved successful in gaining information about the star formation history of galaxies, and are a key way of estimating the stellar mass and star formation rate of galaxies.

1.3.2 Galaxy Formation in Halos

The central paradigm of how galaxies form is that dark matter haloes collapse first, as per theories of structure formation, see section 1.2. Gas then cools in these potential wells, until it is able to start forming stars. Continued star formation, plus galaxies and their host haloes merging, are believed to be key processes in building up the galaxies we see today. See White & Rees (1978); Rees & Ostriker (1977) for early discussions of the processes involved in the initial collapse of gas in haloes. In this section we discuss some of the main observable properties of the galaxy population.

1.3.3 Stellar mass function

The stellar mass function is the comoving number density of galaxies, per unit of stellar mass e.g. a quantification of how many galaxies there are as a function of mass. It is the galaxy equivalent to the halo mass function described in section 1.2.2.2 - figure 1.12 shows an example.

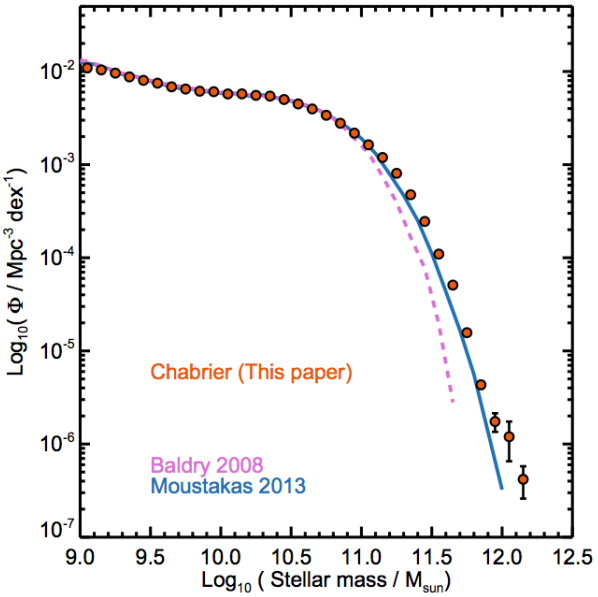


Figure 1.12: The stellar mass function at $z \sim 0.1$ from McGee et al. (2014b).

A closely related statistic is the luminosity function, comoving density as a function of source luminosity in some wavelength. Although closely linked (typically more massive galaxies are more luminous), stellar mass functions and luminosity functions each give subtly different insights into the physics at play. For example, stellar mass can be more directly linked to theory, but luminosities are more robust measurements (e.g. they don't rely on our confidence that our models to estimate stellar mass are correct). Luminosity functions also give different information in some contexts e.g. the radio luminosity function is giving information about the range of AGN activity in the Universe, the UV luminosity function gives information about what star formation is taking place in the Universe etc.

The stellar mass function is observed to have the same qualitative shape as the halo mass function e.g. a power law at low masses, and an exponential cut-off at some critical stellar mass. However, potentially problematically, this gradient of the low mass power law of the stellar mass function is observed to be shallower than that predicted for the halo mass function, and similarly the exponential cut-off at the high mass end is much sharper for the observed stellar mass function than the predicted halo mass function (Silk & Mamon, 2012; Mutch et al., 2013). It is postulated that the complex baryonic physics of galaxy formation is responsible for this deviation, see section 1.3.6.

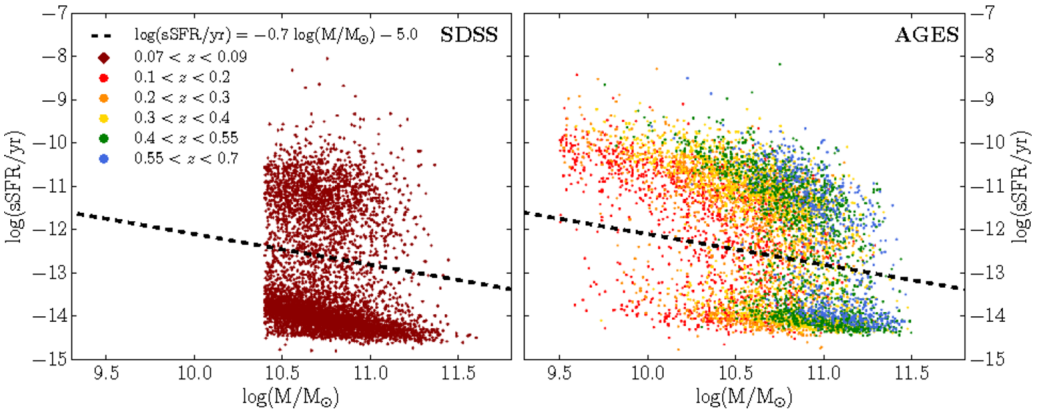


Figure 1.13: The stellar masses and specific star formation rates of galaxies in the local Universe in SDSS, and at low/intermediate redshifts in AGES. The dashed line shows the division between the star forming main sequence and passive galaxies. Plot from Choi et al. (2014).

1.3.4 Red-Blue Bimodality

As shown in figure 1.13 (taken from Choi et al., 2014), when the stellar mass of galaxies is plotted against their star formation rate, galaxies appear to form two separate populations e.g. Bell et al. (2004). These populations are a 'star-forming main sequence' and a 'red and dead' 'passive' population that is forming stars at an extremely slow rate, with a 'green valley' between them. The fraction of galaxies that are passive increases over time, suggesting that galaxies start in the main sequence, at at some point are *quenched* - have their star formation stopped - and enter the passive population. The fact that the galaxies are bimodal, as opposed to lying on a continuum of star formation rates, is suggestive that the transition from one state to another is relatively quick - otherwise we would observe a larger green valley (Schawinski et al., 2014). The lower mass end of the stellar mass function is dominated by star-forming galaxies, and the high mass end by passive galaxies (e.g. higher mass galaxies are forming proportionately fewer stars, Drory et al., 2009). Galaxy star formation rate is also known to be correlated with morphology (spirals have higher star formation rates than ellipticals, e.g. Lee et al., 2013), and both morphology and star formation have a strong environmental dependence (e.g. Dressler, 1980). In general galaxy stellar mass, star formation rate and environment are all highly correlated - untangling how these properties are interrelated, and what causal inferences can be made

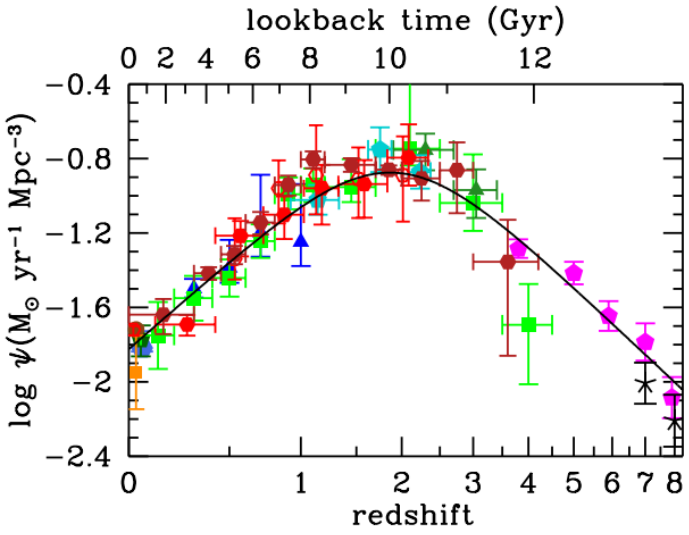


Figure 1.14: Cosmic star formation as a function of redshift. Plot from Madau & Dickinson (2014).

from the correlations, is a major challenge in galaxy evolution, and a key theme of this thesis.

1.3.5 Galaxies Over Cosmic Time

The cosmic star formation rate is the star formation rate integrated over all galaxies, divided by comoving volume e.g. total stellar mass formed per unit of time per unit volume of the Universe. Figure 1.14 from Madau & Dickinson (2014) shows the measured cosmic star formation rate over the history of the Universe.

To some extent the qualitative behaviour of rising and then falling might be expected (if you have a pile of fuel, and set light to it, the energy output will rapidly rise as things start 'burning', before dropping off as the fuel runs out), but the precise shape of this curve is still to be comprehensively understood.

When this qualitative view of the cosmic star formation history first emerged, it became obvious that different probes of the star formation rate gave different normalisations to the relationship, in particular that estimates from the infrared were a lot higher than from UV. Correct accounting of reprocessing by dust is necessary to rectify this discrepancy (e.g. Burgarella et al., 2013).

1.3.6 Active Galactic Nuclei and Feedback Processes

Active Galactic Nuclei (AGN) are ultra-luminous compact components observed at the centre of galaxies, that are believed to be super massive black holes accreting material. As already noted, AGN emit light across the entire electromagnetic spectrum (Risaliti & Elvis, 2004). Despite making-up only a small fraction of the mass of a galaxy ($\sim 0.1\%$ of the galaxy mass), they are believed to potentially be able to shape the evolution of a galaxy dramatically (see Kormendy & Ho, 2013). In particular many authors suggest that AGN activity may cause *feedback* - existing collapsed baryonic matter influencing either positively or negatively the formation of new stars. AGN are postulated to give negative feedback in high mass galaxies i.e. inhibiting star formation in massive galaxies (see Fabian, 2012; Antonuccio-Delogu & Silk, 2010), stopping them growing even more massive - explaining the quicker than HMF drop in the stellar mass function discussed in section 1.3.3. Similarly, supernovae and stellar winds are suggested to inhibit star formation in low mass galaxies, resulting in the extremely shallow low-mass power law behaviour of the stellar mass function (Larson, 1974; Dekel & Silk, 1986; Fierlinger et al., 2016). This general concept of supernovae and AGN feedback shaping the stellar mass function at the low and high mass ends respectively, has had success in making stellar mass functions from SAMs and hydrodynamic simulations agree with observed ones (see Furlong et al., 2015), although it is far from incontrovertible e.g. there is some evidence in the literature that AGN actually give positive feedback in some contexts, stimulating star formation (Silk & Norman, 2009; Silk, 2013; Feain et al., 2007; Zinn et al., 2013).

1.4 Wide Field Surveys

Extragalactic astronomy uses a wide range of observational strategies to observe the cosmic galaxy population. In this thesis I mainly focus on 'wide field surveys', which attempt to observe extremely large number of galaxies in some region of the sky, typically at the expense of detailed information about each individual galaxy.

Wide-field surveys give comparatively little information about each individual object - but give vast numbers of sources for greater statistics alongside large-scale structure infor-

mation. Wide-field surveys bridge the gap between galaxy evolution, large-scale structure and cosmology.

1.4.1 Imaging

Many imaging surveys use a wide-field camera to image large areas (up to several square degrees) of the sky at a time to a certain depth.

In this thesis I use data from the VIRCAM camera on the VISTA telescope (Dalton et al., 2006); other notable wide-field cameras include MegaCam on the Canadian-France-Hawaii Telescope (Boulade et al., 2003), Hyper Suprime-Cam on the Subaru telescope (Miyazaki et al., 2012) and DECam on the Victor M. Blanco Telescope (Mohr et al., 2008).

Surveys, at fixed telescope power and observing time, can either be 'wide' - observe a large area of the sky to low sensitivity (only able to detect brighter objects) - or 'deep' - observe a smaller area of the sky, but to higher sensitivity. See figure 1.15 for a comparison of different survey widths and depths.

1.4.2 Spectroscopic

Spectroscopic surveys (e.g. the VIMOS VLT Deep Survey, Le Fèvre et al., 2013) seek to measure the spectrum of every galaxy in some region of the sky with some property e.g. measure the spectrum of every source in a 1 $\deg^2$ square field brighter than some threshold. Typically it is necessary to have already done an imaging survey over a field to 'discover' the sources before the spectroscopic survey.

Spectroscopic surveys typically have the advantage of precision spectroscopic redshifts, and information from spectral lines e.g. line broadening for AGN, star formation indicators like the [OIII] or Hα lines, an idea of metallicity etc. Imaging surveys typically only have poorer photometric redshifts, but for a larger number of sources. They get the integrated light over large wavelength ranges, so can get a good handle on stellar mass.

In this chapter I have discussed the background and framework of the central topics of this thesis; cosmology, large-scale structure, and galaxy evolution. The rest of this thesis is organised as follows. In Chapter 2 I present various methods for characterising the

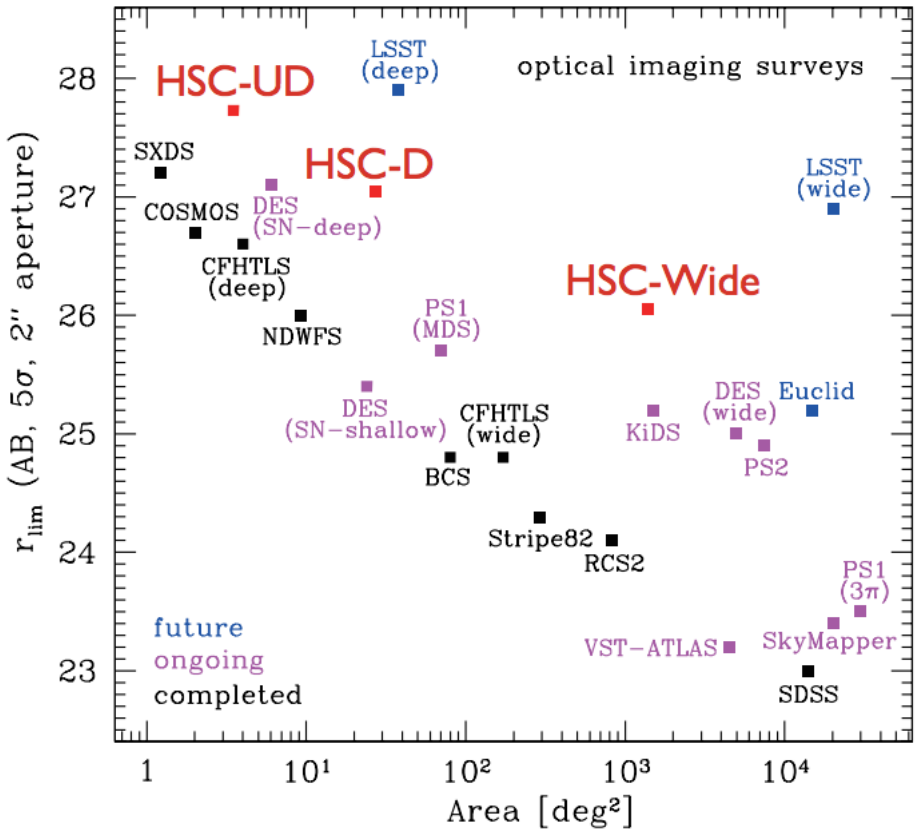


Figure 1.15: Figure from the Hyper Suprime-Cam strategic survey proposal (Miyazaki et al., 2013). Plotted are the areas (in deg^2, x-axis) and depths (AB magnitude, y-axis) of several optical extragalactic surveys. Shown in black are completed surveys, in purple ongoing surveys, in red Hyper-Suprime Cam surveys and in blue future planned surveys (all as of 2012 when the plot was made). Surveys in the top left are extremely deep, but narrow surveys; those in the bottom right are very wide (whole sky $\approx 10^{4.6}$ deg^2), but very shallow. In general, contemporaneous surveys lie on some diagonal - as time goes on and instrumentation improves, the curve of accessible parameter space moves towards the top right. Exceptions can exist e.g. a survey might only be able to achieve a lower depth and area than might be expected if it is also required to have higher resolution. This figure shows optical surveys, but the behaviour of how survey power increases over time is qualitatively the same at other wavelengths.

large-scale distribution of galaxies, in particular the angular correlation function and halo occupation distribution (HOD) models. In Chapter 3 I measure the clustering of galaxies over cosmic time in the VIDEO survey, model these measurements using HOD methods, and derive stellar mass to halo mass ratios. In Chapter 4 I discuss an approach to introduce more physical environmental effects into HOD models, and apply this method to measurements in VIDEO, discussing the implications of the results for our understanding of environmental

quenching. In Chapter 5 I measure the clustering of a sample of very bright Lyman-break galaxies at $z \sim 6$, and explore the consequences for our understanding the 'onset of quenching' in the early Universe. In Chapter 6 I compare clustering measurements in a mock sky from the hydrodynamic cosmological simulation Horizon-AGN to corresponding measurements in the VIDEO survey. Finally in Chapter 7 I summarise my findings, and discuss future planned future research.

Magnitudes of galaxies are given in the AB system (Oke & Gunn, 1983) throughout this thesis unless otherwise stated. All calculations in this thesis are in the concordance cosmology $\sigma_8 = 0.8$, $\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$ and $H_0 = 70$ km s^{-1}Mpc^{-1} unless otherwise stated.

Chapter 2

Measuring and Modelling Large-Scale Structure

As discussed in Chapter 1, the large-scale structure of the Universe is a topic of great interest; both in and of itself, as a probe of cosmology, and for understanding the evolution of galaxies. In this Chapter I discuss various statistical methods for measuring and modelling the large-scale structure distribution of galaxies, in particular the correlation function and the Halo Occupation Distribution (HOD) model respectively, which I focus on in the science chapters of this thesis.

2.1 The Correlation function

The 'base' statistic that I use to quantify LSS in this thesis is the *spatial correlation function*, defined as *the excess probability of finding galaxies at a given separation than would be expected were the distribution of galaxies uniform.* This is normally considered as a function of (comoving) distance:

$$\mathrm{d}P = \rho(1 + \xi(r))\mathrm{d}V, \tag{2.1}$$

where P is probability, ρ is the galaxy comoving number density, $\xi(r)$ is the correlation function as a function of comoving distance, and V is comoving volume. The correlation

function can be considered as a function of scale only, as opposed to a function of location in space and direction, because of the homogeneity and isotropy of the Universe. $\xi(r) > -1$ and $\lim_{r\to\infty} \xi(r) = 0$ for non-negative probabilities and for non-infinite densities respectively.

In this section I discuss variations on this statistic, and how it can be calculated.

2.1.1 Spatial, Projected and Angular Correlation Functions

The correlation function can be measured in 3D space when distances to galaxies are well known. However the fact that often distances are only available from redshifts, if at all, means that it is common to also consider various distorted or projected versions of the statistic.

2.1.1.1 The Redshift-Space Distorted Spatial Correlation Function

Ideally one would have access to a 'birds-eye view' of slices of spacetime in the Universe, and measure the correlation function from this vantage point. However unfortunately we are Earthbound and must make do with whatever observations are possible from Earth. To get the 3D locations of galaxies, we can get two dimensions from angular location on the sky, and, as discussed in section 1.1.3.1, given a knowledge of cosmology, can convert a measured redshift into a cosmological distance. However galaxies also have peculiar motion relative to the Hubble flow, which also imparts a redshift or blueshift, meaning that their redshift is actually

$$1 + z_{obs} = (1 + z_{cosmo})(1 + z_{pec}), \tag{2.2}$$

where z_{obs} is the redshift observed, z_{cosmo} is the redshift that comes from cosmic expansion, and z_{pec} is the blue/redshift from the peculiar motion of the source. Essentially, for two galaxies at the same cosmological distance, both will have the same cosmological redshift, but the one moving towards us will have a slightly lowered redshift, and one moving away will have a slightly higher redshift.

At high redshift, this is not particularly problematic if one is only concerned with the general epoch of the galaxy, as typically $z_{cosmo} >> z_{pec}$, so it doesn't impact on measure-

ments of the stellar mass function at a given redshift etc. However these deviations *do* add up when measuring large-scale structure using redshift as a distance measure, as the effect of misinterpreting the peculiar redshift as a cosmological redshift is effectively stacked over all galaxies. This was first considered by Jackson (1972), and further developed in Kaiser (1987). On small scales the impact of peculiar velocities is illustrated by considering galaxies moving around in a potential well with some velocity dispersion. They will all have essentially the same cosmological redshift, but a range of peculiar redshifts (depending on the velocity dispersion). Directly converting the measured redshift into a cosmological distance will elongate the structure along the line of sight of the observer (the 'fingers of god' effect; Fig. 2.1), but no such elongation will occur in the transverse direction. Ignoring these redshift-space distortions would present a Universe with all structure 'pointing' towards us - which of course the Copernican principle should lead us to be suspicious of. On larger scales the opposite effect happens. Matter is pulled towards an over density from all sides. Matter on the other side of the over density to the observer is pulled towards it, and is slightly blueshifted. Matter on the observer side of the overdensity is also pulled towards it, and is slightly redshifted. The net effect of this is that structures on large scales are flattened ('pancakes of god').

The fingers and pancakes of god effects show up when measuring the correlation function in galaxy surveys. The spatial correlation function now becomes the *anisotropic* spatial correlation function $\xi(r_p, \pi)$, where r_p is the projected distance perpendicular to the line of sight, and π is the distance along the line of sight. Figure 2.2 shows the impact of the galaxy peculiar velocities on the correlation function.

The effects of galaxy peculiar velocities and coherent infall are both essentially special relativity effects. However there are also expected to be general relativistic effects. One of the most simple of which is gravitational redshift distortion, which is effectively (phrasing a relativistic concept in common language) photons being redshifted when they expend energy leaving a potential well. This gives sources slightly higher redshifts than they would have from just cosmological redshift and their peculiar motion. Furthermore, massive galaxies are typically in deeper gravitational wells than lower mass galaxies, so massive galaxies will appear to be further away than they actually are compared to lower mass galaxies. This

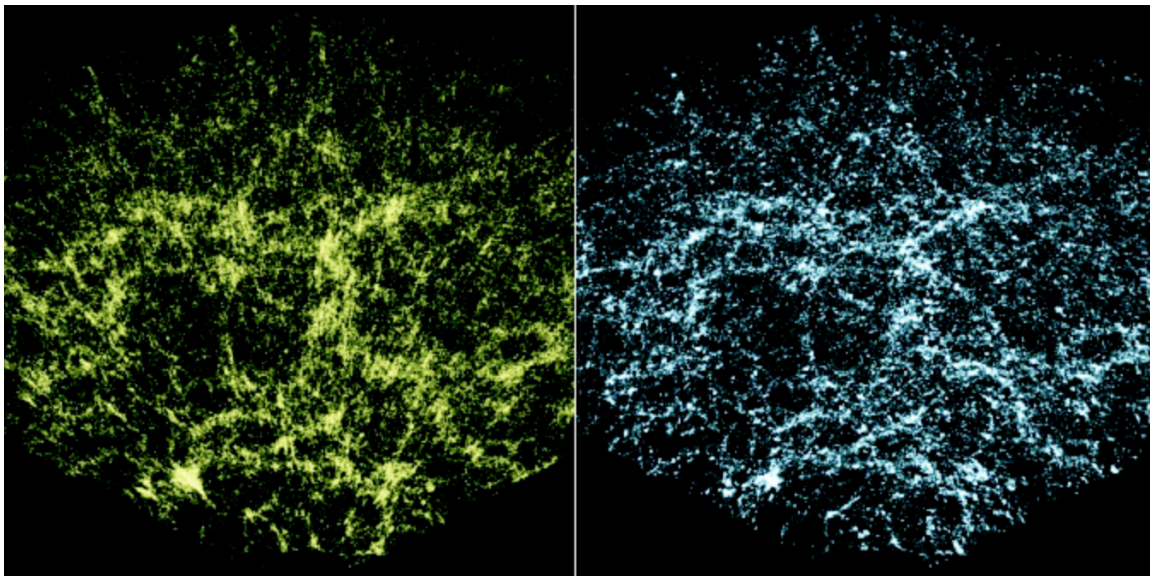


Figure 2.1: Visualisation of the Fingers-of-God effect in the Sloan Digital Sky Survey from SubbaRao et al. (2008). The observer is in the centre at the bottom of each figure. The left half of the figure shows large-scale structure converting redshifts to distances ignoring peculiar motion. The right half of the figure shows the same data with redshift-space distortions 'corrected', using an approach from Tegmark et al. (2004).

gives an asymmetric effect when massive and less massive galaxies are cross-correlated, see figure 2.2 - essentially this makes the lower mass galaxies be more likely to appear to be on the 'near' side of the massive galaxy. Since the first detection of gravitational effects on large scale structure in Wojtak et al. (2011) the field is still in its infancy, although as Croft (2013) note there are many upcoming surveys where it may be possible to measure such an effect, e.g. the eBOSS survey (Dawson et al., 2016).

2.1.1.2 Integrating to the Projected Correlation Function

Redshift space distortions contain a lot of information about the dynamics of large-scale structure. However it is common to integrate out these distortions if one is purely interested in the isotropic correlation function. This is known as the *projected correlation function*,

$$w(r_p) = 2\int_0^\infty \mathrm{d}\pi \xi\left(\sqrt{r_p^2+\pi^2}\right) = 2\int_{r_p}^\infty \mathrm{d}r \frac{\xi(r)}{\sqrt{r^2-r_p^2}}, \tag{2.3}$$

where $w(r_p)$ is the projected correlation function, and r_p is the projected distance be-

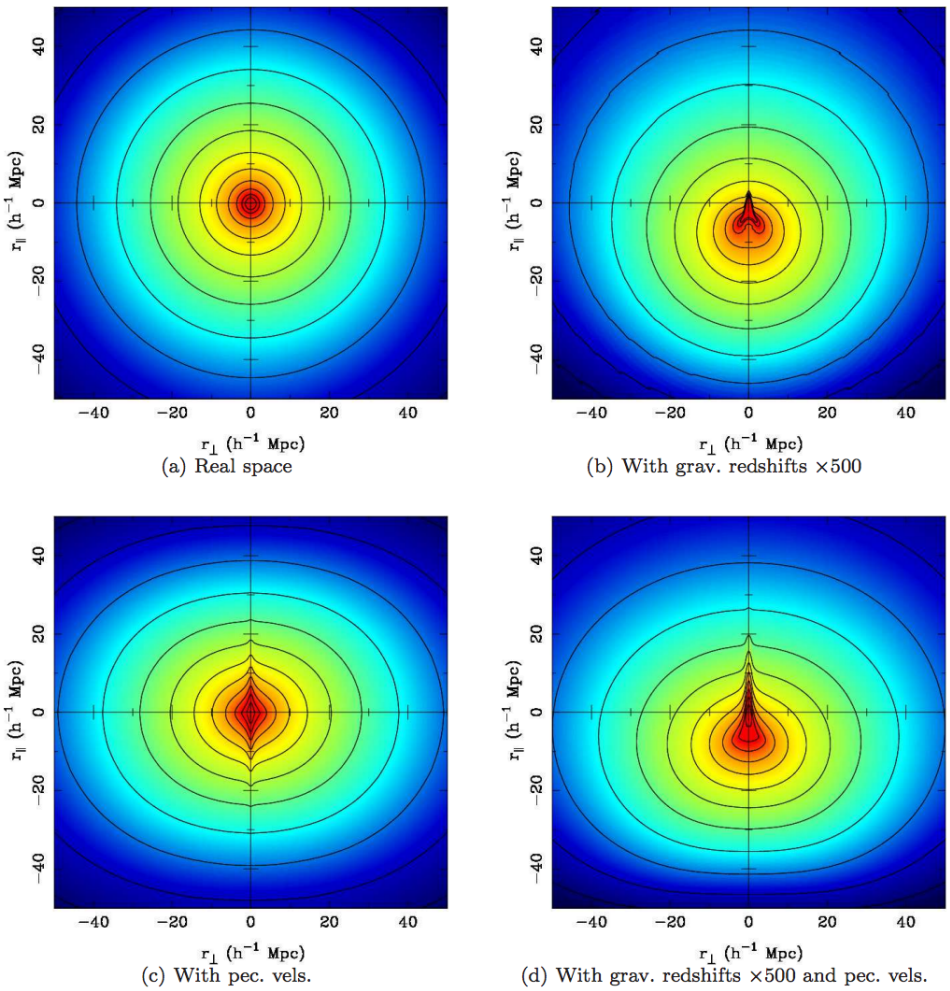


Figure 2.2: Toy-model illustration of redshift-space distortion of the cross correlation of massive to less massive galaxies from Croft (2013). Top left: contours of an isotropic spatial correlation function. Top right: gravitational redshift added. Bottom left: peculiar motion added. Bottom right: gravitational redshift and peculiar motion added. Note that the size of the gravitational redshift has been multiplied by 500 for it to be visible - it is a much smaller effect than peculiar velocities.

tween sources. The projected correlation function essentially removes one dimension from the correlation function, and measures how much more likely than Poisson galaxies are to be at a given projected distance from one another.

2.1.1.3 Projecting to the Angular Correlation Function

Often only photometric redshifts are availible (see Chapter 3), which are accurate enough to place a galaxy in the correct epoch, but not accurate enough for small-scale structure to

be measured to high precision. In these cases, it can be convenient to consider the angular correlation function, the excess probability above Poisson of finding two galaxies at a given angular separation:

$$dP = \sigma(1 + \omega(\theta))d\Omega, \tag{2.4}$$

where ω is the angular correlation function, σ is the galaxy angular surface density, θ is angular separation and Ω is solid angle.

The angular correlation function can be calculated from the spatial correlation function via:

$$\omega(\theta) = 2\int_0^\infty \mathrm{d}x\, f(x)^2 \int_0^\infty \xi(\sqrt{u^2 + (x\theta)^2})\mathrm{d}u, \tag{2.5}$$

where x and u are comoving distance (down the line of sight), f(x) is the comoving distance distribution (derived from the redshift distribution), θ is the angular scale and ξ is the spatial correlation function (see Limber, 1954; Parry, 1977). The projection assumes the small angle approximation and that the spatial correlation function does not evolve dramatically in this redshift range.

In this thesis, I typically consider some model spatial correlation function, and then project it to recover the model angular correlation function, to compare with a measured angular correlation function.

2.1.1.4 Transforming to the Power Spectrum

Alternatively the distribution of matter and galaxies can be described in Fourier space by the power spectrum $P(k)$, which can be converted to and from the spatial correlation function via:

$$\xi(r) = \frac{1}{2\pi^2}\int \mathrm{d}k\, k^2 P(k)\frac{\sin(kr)}{kr}, \tag{2.6}$$

and

$$P(k) = \frac{1}{2\pi^2} \int \mathrm{d}r\, r^2 \xi(r) \frac{\sin(kr)}{kr}. \tag{2.7}$$

This is the same statistic discussed in section 1.2.1. The interpretation of the value of the power spectrum for a given value of k is the amplitude of modes of wavenumber k.

Figure 2.3 shows the spatial, projected and angular correlation functions, and the corresponding power spectrum, for the same underlying galaxy statistics.

2.1.2 Calculating the Angular Correlation Function

In this section, I describe methods of measuring the angular correlation function in galaxy surveys, although the equivalent methods in the context of the spatial correlation function are the same.

The most common way to estimate $\omega(\theta)$ is to compare the distribution of the real sources to a 'dummy' catalogue of randomly distributed points constructed to have the same properties as the real data, apart from spatial distribution. This typically means applying the same mask to the dummy data, giving the dummy data the same redshift distribution as the real data for the anisotropic correlation function etc. Then the number of galaxy pairs of separation θ, $DD(\theta)$, is calculated, normalised by the total number of pairs ($n[n-1]/2$, where n is the number of galaxies). Then one also calculates $RR(\theta)$ for the random galaxies.

To determine the angular correlation function, I use the Landy & Szalay (1993) estimator:

$$\omega(\theta) = \frac{DD - 2DR + RR}{RR}, \tag{2.8}$$

which also uses $DR(\theta)$, data to random pairs, as it has a lower variance (as an estimator) and takes better account of edge effects. Other estimators are discussed and compared in Kerscher et al. (2000).

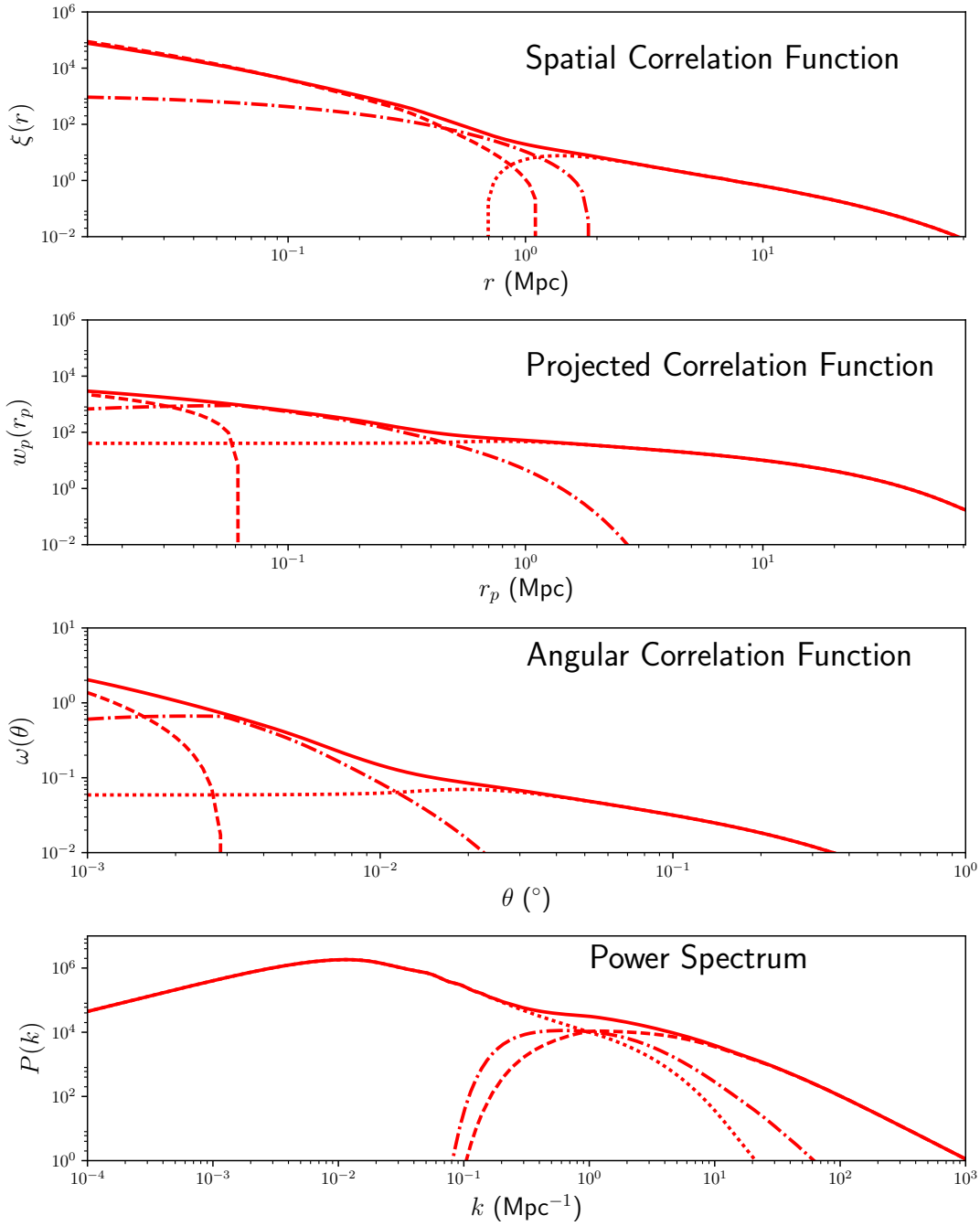


Figure 2.3: Illustration of the same galaxy 2-point statistics through a) the spatial correlation function (first row), b) projected correlation function (second row), c) angular correlation function (third row) and d) power spectrum (bottom row). In each sub-plot the full line shows the full function, dotted line the 2-halo term, dashed line the central-satellite part of the 1-halo term, and dot-dashed line the satellite-satellite part of the 1-halo term. The 2-halo term goes to zero on small scales because we do not allow haloes to overlap, pure linear theory would extend it to arbitrarily small scales. The model correlation function here is from a Zheng et al. (2005) HOD model (see section 2.3, for $z = 1$, and $M_{min} = 10^{12} M_\odot$, $M_1 = 10^{13} M_\odot$, $\alpha = 1$, $\sigma = 0.2$ and $M_0 = 10^8 M_\odot$).

2.1.2.1 Error Bars

By averaging over multiple dummy data sets and using $\overline{RR(\theta)}$, or by letting the number of random data points go to infinity, the error in $RR(\theta)$ can be considered zero i.e. essentially becomes a function of the field geometry. This leaves $DD(\theta)$ as the main source of variance in the estimation, and is often given as the Poisson error in the DD counts:

$$\Delta\omega = \frac{1 + \omega(\theta)}{\sqrt{DD}}. \tag{2.9}$$

However this naive approach can significantly underestimate the uncertainty because adjacent DD bins are correlated. More rigorous approaches therefore rely on bootstrap methods. The 'jack-knife' method consists of blocking off segments of the field and recalculating to see how much the estimate of the correlation functions changes (e.g. Weinberg et al., 2004). 'Bootstrap resampling' consists of sampling the galaxies with replacement from the dataset and recalculating (see Ling et al., 1986). Repetition of this process allows the variance of the $\omega(\theta)$ values to be estimated. Cress et al. (1996) and Lindsay et al. (2014) found Poisson errors were a factor of 1.5 to 2 smaller than those estimated with bootstrap - see Norberg et al. (2009) for a more in depth discussion.

2.1.2.2 The Integral Constraint

When using a finite field, this estimate of the correlation function has to average to 0 (because DD and RR are normalised to sum to the same value). This effectively means that on large scales the number of data pairs is slightly underestimated, which biases this estimate of the correlation function to lower values. The correlation function essentially receives a small negative offset, which is known as the *integral constraint*:

$$\omega_{obs}(\theta) = \omega_{true}(\theta) - K_{IC}, \tag{2.10}$$

where $\omega_{obs}(\theta)$ is the correlation function naively calculated from equation 2.8, $\omega_{true}(\theta)$ is the true correlation function, and K_{IC} is a constant, the integral constraint.

K_{IC} has an analytic expression from Groth & Peebles (1977):

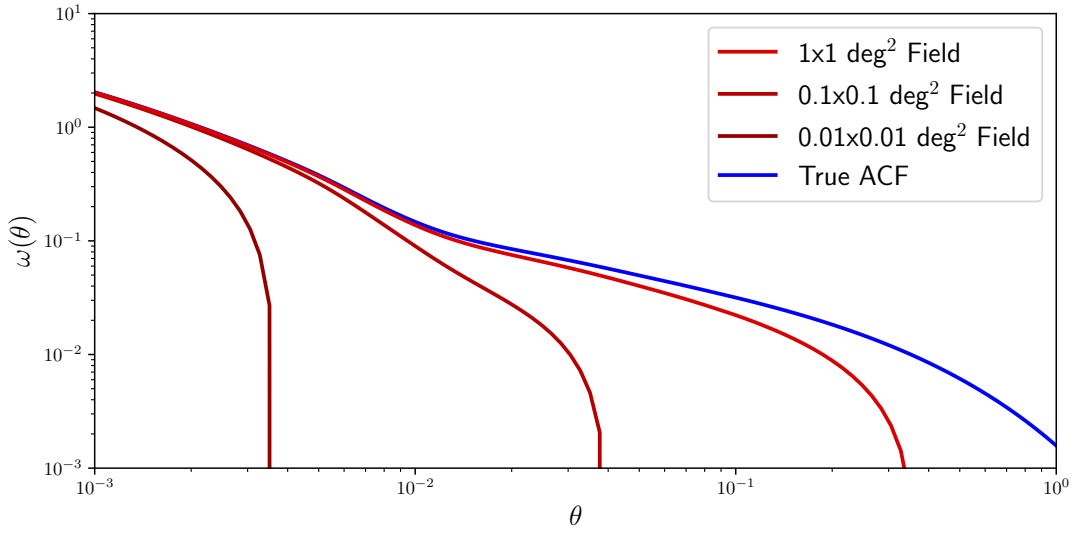


Figure 2.4: A model angular correlation function, and the impact of the integral constraint on the measured angular correlation function when it is measured in three square fields of different sizes. The angular correlation function can be measured accurately on scales much smaller than the size of the field, but is heavily biased downwards on scales comparable to the field size. The model correlation function here is from Zheng et al. (2005) (see section 2.3, for $z = 1$, and $M_{min} = 10^{12} M_{\odot}$, $M_1 = 10^{13} M_{\odot}$, $\alpha = 1$, $\sigma = 0.2$ and $M_0 = 10^{8} M_{\odot}$.)

$$K_{IC} = \frac{1}{\Omega^2} \iint \omega_{true}(\theta) d\Omega_1 d\Omega_2, \tag{2.11}$$

where Ω is the angular area of the field and $d\Omega_1 d\Omega_2$ denotes integrating twice over the field solid angle, which can be estimated numerically (Roche & Eales, 1999) using:

$$K_{IC} = \frac{\Sigma RR(\theta)\omega_{true}(\theta)}{\Sigma RR(\theta)}. \tag{2.12}$$

Some authors (e.g. Coupon et al., 2012) prefer to estimate the integral constraint by modelling their estimate of the correlation function using a power law, estimating the integral constraint from that model, and then adding that value onto their correlation function data points to get 'corrected' points. In this thesis I take the opposite approach of letting the integral constraint be a function of the model, and taking the value off the model to compare to the data, as in Beutler et al. (2011), although there is comparatively little difference between the approaches.

The integral constraint has the effect of reducing the measured correlation function

at large angles and steepening the gradient. Figure 2.4 shows the impact of the integral constraint for different field sizes.

2.1.2.3 Continuous Estimation of the Correlation Function

Approaches to calculating the correlation function conventionally involve binning; the galaxy angular separations are put into angular distance bins (often spaced logarithmically). Although advantageous in terms of simplicity to calculate, and clearer interpretation, binning data is non-ideal in the sense that it i) loses information and ii) can involve arbitrary choice of bin size. In this thesis (see also Hatfield et al., 2016) I adopt an alternative estimator that finds the correlation function as a continuous function, which I describe here.

The approach I use is to implement a non-parametric method for the estimation of $DD(\theta)$, $DR(\theta)$ and $RR(\theta)$, and then use the estimator of Landy & Szalay as per usual. I use the kernel based density estimator (Parzen, 1962; Rosenblatt, 1956), on the set of angular separations to find DD, DR and RR separately, and then choose a kernel bandwidth to minimise the mean integrated squared error (MISE) for each. Heuristically the process can be described thus: first the $\frac{1}{2}N^2$ galaxy separations are calculated. However rather than being binned by angular separation, the distribution is calculated by summing kernel distributions (e.g. normal, top hat or tricube etc.) centred on each point, and kernel width replaces the role of bin size. If the width of the kernel h is too large, the data are over smoothed, and features are lost. If the width is too small, the data is too noisy. There exists an optimal choice that minimises the expected error on this method as an estimator of the true distribution.

I give a brief description of how to choose optimal smoothing parameters as described in Parzen (1962). Suppose $f(x)$ is the true function that one is attempting to estimate (in this context it could be DD) and that $\hat{f}(x)$ is the estimation of the function from the data. The quantity to be minimised is the expected error accumulated over all x, the MISE:

$$\mathrm{MISE} = \mathbb{E}\left(\int (f(x) - \hat{f}(x))^2 \mathrm{d}x\right), \tag{2.13}$$

which can be re-arranged to:

$$\text{MISE} = \int b^2(x)\mathrm{d}x + \int v(x)\mathrm{d}x, \tag{2.14}$$

where

$$b(x) = \mathbb{E}(\hat{f}(x)) - f(x), \tag{2.15}$$

the bias of the estimator at a point and

$$v(x) = \mathbb{V}(\hat{f}(x)), \tag{2.16}$$

the variance of the estimator at a point. Hence MISE is a function of the data and the smoothing parameter h. Minimising MISE is a compromise between minimising bias and minimising estimator variance.

If the data are X_i (in this case galaxy separations) and the kernel is K (a smooth symmetric function around zero that integrates to 1 and goes to zero sufficiently fast; e.g. I use a Gaussian), then the estimate of the function becomes:

$$\hat{f}(x) = \frac{1}{nh}\sum_i K\left(\frac{x - X_i}{h}\right), \tag{2.17}$$

where h can be chosen to be the standard deviation of the kernel. Parzen (1962) show that this estimate is consistent and that the MISE goes as:

$$\text{MISE} \approx \frac{1}{4}c_1^2 h^4 \int (f''(x))^2\mathrm{d}x + \frac{\int K^2(x)\mathrm{d}x}{nh}, \tag{2.18}$$

and the optimal h (by differentiation) is

$$h_* = \left(\frac{c_2}{c_1^2 A(f) n}\right)^{1/5}, \tag{2.19}$$

where $c_1 = \int x^2 K(x)\mathrm{d}x$, $c_2 = \int K(x)^2\mathrm{d}x$ and $A(f) = \int (f''(x))^2\mathrm{d}x$.

Calculating the optimal h value in this manner formally depends on knowing the true distribution (e.g. if f oscillates wildly with high frequency $A(f)$ is high, necessitating a smaller h to pick out features). I calculate this heuristically by doing a first run with a trial

value of h, fitting a power law to the resulting correlation function, and using this as the true distribution for the purposes of finding h. I then subsequently use the estimate of the correlation function that results as my estimate of the true value. In this thesis I typically find $h \approx 0.1$ dex in angular separation, of comparable order to bin sizes most authors choose heuristically.

There do exist entirely data driven cross-validation techniques of choosing h optimally that I do not discuss here (see Bowman, 1984 for a discussion). For continuous estimation of the error (which now takes the form of a band along the estimation of the function) I repeat the discussed process on bootstrapped data sets and take the 16th and 84th percentiles of the multiple estimations of the correlation function.

To confirm that this is consistent with the binning method, I calculated the correlation function with the binning approach as well as the kernel approach for one of the samples discussed in Chapter 3 with $1.25 < z < 1.5$, $M_\star > 10^{10.6} M_\odot$ (Fig. 2.5). A way of viewing the kernel approach is that it is essentially the same as binning - except binning uses a top hat kernel of arbitrary size for each data point, and does not always place the kernel directly on top of the data.

2.1.2.4 Accounting for Redshift Probability Distribution Functions

Often it will be of interest to measure the angular correlation function at different redshifts. However, particularly for photometric surveys, there can be substantial uncertainty on the redshift. The most straightforward approach is to just use the best estimate of the redshift value for the source, without consideration of the uncertainty in the measurement e.g for sources with broad redshift probability density functions (pdfs) the chance of the object being in a different redshift bin to its best fit is not accounted for. In this thesis I take the approach of Arnouts et al. (2002) (shown in Asorey et al., 2016 to reduce bias), and for each redshift bin I assign galaxies a weight corresponding to the probability of the galaxy being in that redshift range according the the photometric redshift PDFs ($p_i(z)$), e.g.

$$W_i = \int_{z_{lower}}^{z_{upper}} p_i(z)\, \mathrm{d}z, \tag{2.20}$$

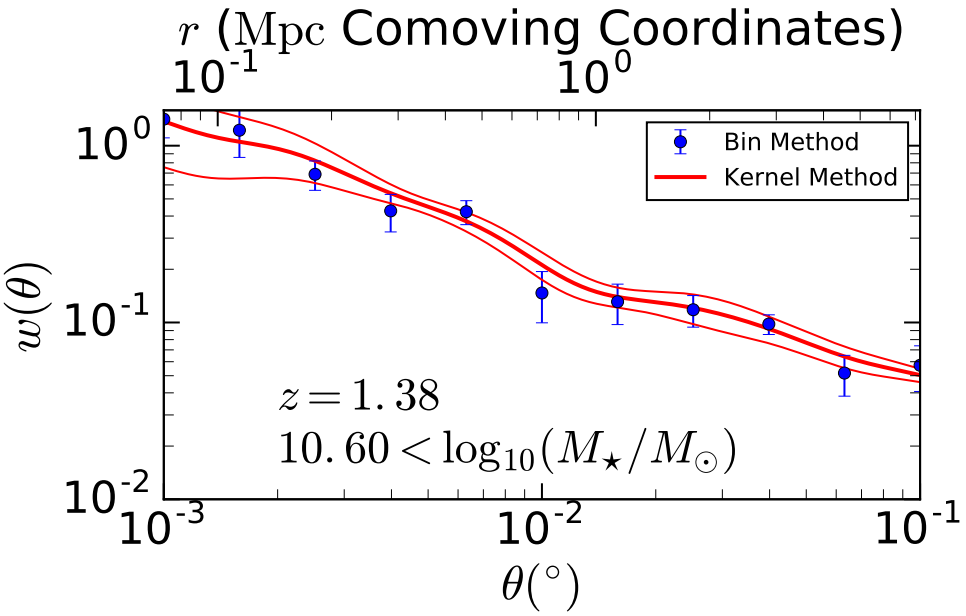


Figure 2.5: Illustration of the agreement of the binning approach and the non-parametric approach to correlation function calculation for a sample with $1.25 < z < 1.5$, $M_\star > 10^{10.6} M_\odot$. The error bars (the secondary lines in the case of the kernel method) represent the 16th and 84th percentiles from bootstrapping. Full details of the measurements can be found in Chapter 3.

and

$$DD(\theta) = \frac{2}{n(n-1)h} \sum_{i,j} W_i W_j K \left(\frac{\theta - d(G_i, G_j)}{h} \right), \tag{2.21}$$

where K is the chosen kernel, h is the kernel width, $d(G_i, G_j)$ is the angular separation of galaxies i and j, and n is the number of galaxies, $n(n-1)/2$ being the number of galaxy pairs. In the limit of highly accurate redshifts, this method reduces to the approach of just working with galaxies where the probability density function has its peak in the bin.

An alternative way of accounting for uncertainty in redshift estimates is to calculate the correlation function just taking the peak of the pdf as the redshift value, but then for the deprojection use a redshift distribution that is the sum of all the pdfs of the sources with their peak in that redshift range, as per Coupon et al. (2012) and Lindsay et al. (2014).

2.2 Abundance Matching and Power Law Modelling of the Correlation Function

As discussed in section 1.3, dark matter haloes are believed to have a strong influence on determining the properties of their constituent galaxies (e.g. Kauffmann et al., 1997; Benson et al., 1999; Norberg et al., 2002; Lewis et al., 2002; Hogg et al., 2003; Croton & Darren, 2007). However, the dark matter halo is invisible - obtaining information about the host dark matter haloes of galaxies is a key theme of this thesis. In this section I highlight two methods for using large-scale structure to learn about the host dark matter haloes of galaxies, that are in some sense forerunners to the more sophisticated HOD method which I discuss in section 2.3.

2.2.1 Abundance Matching

Abundance matching works on the principle that the most massive(/luminous) galaxies are in the most massive haloes. When the theoretical halo mass function is taken with a measured stellar mass(/luminosity) function, then galaxies can be paired up with haloes to give stellar mass(/luminosity) as a function of halo mass:

$$M_h(M_\star) = \psi^{-1}_{HMF}(\psi_{lum}(M_\star)),$$

where M_h is the halo mass that hosts galaxies of mass $M_\star$, $\psi_{HMF}(M_h)$ is the comoving number density of haloes greater than mass M_h and $\psi_\star(M_\star)$ is the comoving number density of galaxies brighter than magnitude $M_\star$. More sophisticated implementations match galaxy groups (that have been pre-identified) to haloes, see Yang et al. (2007). In Chapter 5 I present an application of abundance matching and a more in depth discussion of when abundance matching fails.

2.2.2 Power Law Modelling of the Correlation Function

Galaxy correlation functions are measured to be close to power laws in most galaxy surveys, with an index ~ -1.8 for the spatial correlation function, and ~ -0.8 for the projected and

angular correlation functions. If one fits a power law to a measured correlation function, the bias (see section 1.2.2.3) of the galaxy population can be calculated via $b = \sqrt{\frac{\xi_{gal}(r)}{\xi_{DM}(r)}}$. This bias can then be matched to the halo mass with that bias to get an estimate of the mass of the halo expected to host those galaxies. Power law modelling however only gives an indication of the *average* halo mass - halo occupation modelling (section 2.3) is required to gain insight into the *distribution* of host halo masses.

2.3 Halo Occupation Distribution Modelling

Halo Occupation Distribution (HOD, see Seljak, 2000; Peacock & Smith, 2000; Berlind et al., 2003; Zheng et al., 2005; Zehavi et al., 2005 for early work on the model) descriptions of correlation functions have seen great success in recent years in modelling the galaxy correlation function to high degrees of precision, and are found to give physical results in good agreement with other methods (e.g. Simon et al., 2009, Coupon et al., 2015). The central premise of HOD modelling is to divide clustering into '2-halo' clustering, pairs of galaxies in different haloes, that follow the large scale linear matter fluctuations, and '1-halo' clustering, the highly non-linear clustering of galaxies within individual haloes. Models typically prescribe the mean number of galaxies in a halo as a function of halo mass, assume the occupation number has a Poissonian distribution, and assume that these galaxies trace the halo profile. Then the HOD model, choice of halo profile, halo mass function and a bias proscription can be translated into a spatial correlation function, and then projected to an angular correlation function. Parametrising the HOD allows physical information to be extracted via fitting the model angular correlation functions to measured correlation functions. Variants include fitting simultaneously with cosmology (van den Bosch et al., 2013), varying the compactness of the profile the galaxies follow, allowing the occupation statistics to be non-Poissonian and investigating if different galaxy samples occupy the halos independently (Simon et al., 2009). It is also possible to fit HOD models using galaxy-galaxy weak lensing data (e.g. Coupon et al., 2015), background counts in halo catalogues (e.g. Rodriguez et al., 2015), and even abundance matching techniques (e.g. Guo et al., 2016). The modelling approach I use in this thesis follows closely that of Coupon et al. (2012) and

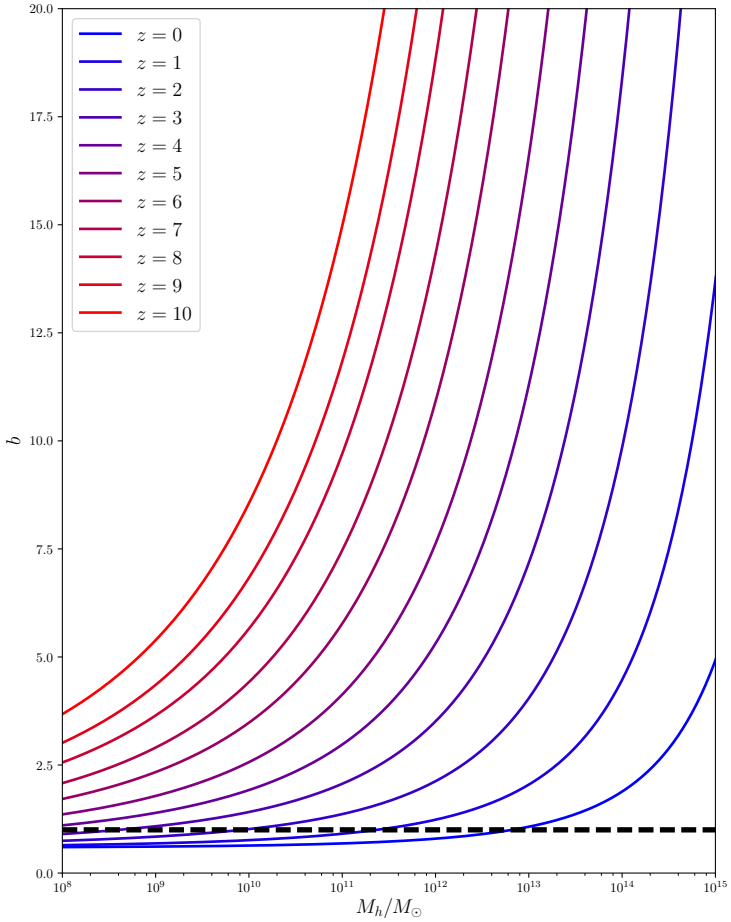


Figure 2.6: Illustration of how the bias of halos of a given mass evolves with redshift. The dashed line shows $b = 1$. Halo bias increases rapidly with halo mass and redshift. By $z = 0$, bias is close to one, and slightly less than one for lower mass haloes. Halo bias taken from Tinker et al. (2010).

McCracken et al. (2015).

The redshift dependence of the various 'ingredients' of the HOD model must be taken into account; Figures 2.6 and 2.7 show how halo bias and the halo mass function evolve with redshift, see section 1.2.2.

2.3.1 Physics Probed

HOD modelling treats the number of galaxies (of a given type) in a halo as a random variable that is a function of halo mass. The purpose of HOD modelling is to learn more about this random variable.

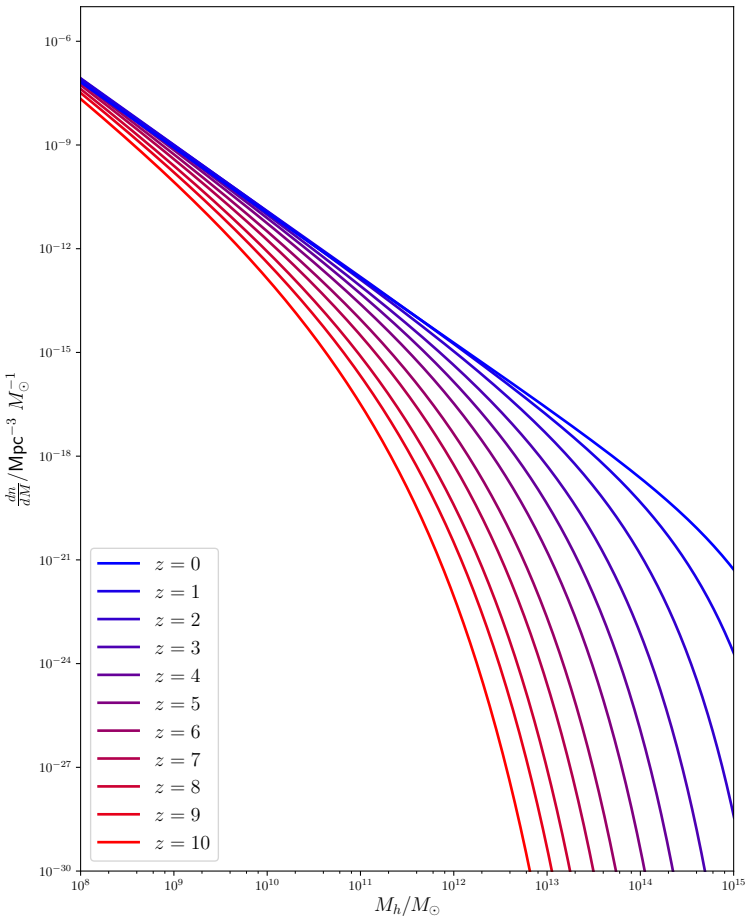


Figure 2.7: The halo mass function (comoving number density per unit of mass) as a function of halo mass, for a range of redshifts. The HMF increases with time (/decreases with redshift), and decreases with increasing halo mass. The characteristic power law with high mass exponential cut-off behaviour discussed in section 1.2.2.2 is evident. By $z = 0$ the HMF is essentially a power law over all halo masses of interest. Halo mass function taken from Tinker et al. (2010).

Most models treat central and satellite galaxies separately, and define $N_{cen}(M_h)$, the random variable for the number of central galaxies as a function of halo mass (which can normally only take the values zero or one), and $N_{sat}(M_h)$, the random variable for the number of satellite galaxies in the halo as a function of halo mass (which takes non-negative integer values). This leads naturally to $N_{tot}(M_h)$, the total number of galaxies in a halo:

$$N_{tot} = N_{cen} + N_{sat}. \tag{2.22}$$

2.3.2 Derived Parameters

Given a HOD model, one can derive several basic commonly measured properties of a galaxy population.

The comoving number density of a population is the number of members of that population per unit comoving volume. This is defined for centrals by:

$$n_{cen} = \int_0^\infty \langle N_{\rm cen}(M_h)\rangle HMF(M_h) {\rm d}M_h, \tag{2.23}$$

and similarly for n_{sat} and n_{tot} (and of course $n_{tot} = n_{sat} + n_{cen}$).

We can calculate the fraction of galaxies that are satellites as

$$f_{sat} = \frac{n_{sat}}{n_{tot}} \tag{2.24}$$

The central fraction can be defined as $f_{cen} = 1 - f_{sat}$.

The effective bias of the galaxy sample is:

$$b_{\rm eff} = \frac{1}{n_{tot}} \int_0^\infty \langle N_{\rm tot}(M_h)\rangle HMF(M_h) b(M_h) {\rm d}M_h \tag{2.25}$$

e.g. the averaged halo bias, weighted by the number of galaxies in those haloes.

2.3.3 The 2-halo Term

The model of the correlation function on large scales is essentially the matter correlation function scaled by the effective bias squared e.g.

$$\xi_{gal}(r) = b_{\rm eff}^2 \xi_{DM}(r). \tag{2.26}$$

A slightly refined version that is normally used is:

$$P_{\rm 2,gal} = P_{DM}(k) \times \left[\frac{1}{n_{tot}} \int_0^\infty \langle N_{\rm tot}(M_h)\rangle HMF(M_h) b(M_h) |u(k,M)| {\rm d}M_h \right]^2, \tag{2.27}$$

which can then be translated back to real space as per section 2.1.1.4, where $u(k, M)$ is the

Fourier transform of the halo profile as a function of halo mass. This correction is designed to account for the fact that halos are extended objects and not all galaxies are at the centre of the halo, so this expression gives a small boost on scales larger than halos, but smaller than large-scale linear perturbations. This corresponds to galaxy pairs in halos that are nearby, where the galaxies are on the side of the halo near to the other halo. On large scales (small k values), $u(k, M)$ tends to 1, and the simpler expression is recovered.

2.3.4 The 1-halo Term

The contribution from galaxies within a single halo to the small scale correlation function is the set of pairs of separations of galaxies within the halo. This is proportional to the convolution of the halo profile with itself[1]. It is within the 1-halo term that the distinction between central and satellite galaxies comes into play. The central galaxy is assumed to be found at the centre of the halo, and the satellites trace the halo profile. Thus the 1-halo term can be further broken down to a 'central-satellite' term, of central-to-satellite pairs, and a 'satellite-satellite' term, of satellite-to-satellite pairs.

2.3.4.1 The Central-Satellite Term

The central galaxy essentially has a 'profile' of a Dirac-δ function at the centre of the halo. The central-satellite term is a convolution of this term and the halo profile. The convolution of a point/Dirac-δ and a profile is just the profile, thus the 1-halo central-satellite term is:

$$1 + \xi_{cs}(r) = \frac{1}{n_{tot}^2/2} \int_0^\infty \langle N_{\rm sat} N_{\rm cen}(M_h) \rangle HMF(M_h) \rho(r, M_h) {\rm d}M_h \tag{2.28}$$

where $\rho(r, M_h)$ is the profile of the halo as a function of halo mass (normalised to have total weight 1), and $\langle N_{\rm sat} N_{\rm cen}(M_h) \rangle$ is the expected number of satellite-central pairs. The $n_{tot}^2/2$ is a normalisation for total number of pairs.

[1] In this thesis I will typically use the word 'convolution' when referring to auto/cross-correlating density profiles, and only use auto/cross-correlation to refer to the two-point galaxy auto/cross-correlation function. Since the distributions are spherically symmetric, convolutions and auto-correlations coincide.

2.3.4.2 The Satellite-Satellite Term

The satellite-satellite term is essentially the weighted averaged halo profiles convolved with themselves. The weight for each halo mass is the number of satellite-satellite pairs,

$$1 + \xi_{ss}(r) = \frac{1}{n_{tot}^2/2} \int_0^\infty \frac{\langle N_{\rm sat}(N_{\rm sat} - 1)\rangle}{2} HMF(M_h)(\rho \star \rho)(r, M_h) \mathrm{d}M_h. \tag{2.29}$$

Where $(\rho \star \rho)(r)$ is the convolution of the halo profile with itself, and $\frac{\langle N_{\rm sat}(N_{\rm sat}-1)\rangle}{2}$ is the expected number of satellite-satellite pairs This is more easily calculated in Fourier space:

$$P_{ss}(k) = \frac{1}{n_{tot}^2/2} \int_0^\infty \frac{\langle N_{\rm sat}(N_{\rm sat} - 1)\rangle}{2} HMF(M_h)|u(k, M)|^2 \mathrm{d}M_h \tag{2.30}$$

2.3.5 Common Parametrisations

The previous subsections describe how to obtain a model correlation function, given random variables N_{cen} and N_{sat}. To compare to data one typically needs to parametrise these random variables, and then fit for the parameters that describe this model.

The basic model often used is that below a critical mass there are no galaxies (of that population) in the halo. Then above that mass, haloes normally have a central galaxy. Then, at a higher mass, halos start to have satellite galaxies, where the number of haloes grows as a power law, and the occupancy distribution is roughly Poisson. In this thesis I use the 5 parameter model of Zheng et al. (2005):

- $M_{\rm min}$, minimum halo mass required for the halo to host a central galaxy
- M_1 the typical halo mass for satellites to start forming
- α as the power law index for how the number of satellites grows with the halo mass
- $\sigma_{\log_{10} M}$ parametrises how discrete the cut off in halo mass for forming a central galaxy is, and
- M_0 a halo mass below which no satellites are formed.

We assume a Navaro-Frenk-White profile (Navarro, Frenk, & White, 1997) with halo compactness parameters from Bullock et al. (2001) and a Tinker et al. (2010)[2] HMF and bias model. We use the HALOMOD python package[3] to calculate the correlation functions.

Quantitatively these halo parameters correspond to:

$$N_{\rm cen} = \begin{cases} 1, & \text{with probability } \frac{1}{2}\left(1 + \text{erf}\left(\frac{\log_{10} M_{\rm halo} - \log_{10} M_{\rm min}}{\sigma_{\log_{10} M}}\right)\right) \\ 0, & \text{otherwise} \end{cases}, \tag{2.31}$$

$$N_{\rm sat} = \begin{cases} 0, & \text{if } N_{\rm cen} = 0 \\ \text{Poiss}\left(\mu = \left(\frac{M_{\rm halo} - M_0}{M_1}\right)^{\alpha}\right), & \text{if } N_{\rm cen} = 1 \end{cases}. \tag{2.32}$$

That is to say haloes have no satellites if they do not have a central galaxy. Note the slight misuse of notation: the satellite power law term should never go negative when $M_h < M_0$, but I show it as the power law for clarity to avoid putting '$\max(0, M_h - M_0)$' every time.

Thus the number of central galaxies as a function of halo mass behaves as a softened step function and the number of satellites is a power law that initiates at a characteristic mass. The equations only allow there to be satellite galaxies when there is a central galaxy.

From these expressions, the various other necessary expectations can be derived:

$$\langle N_{\rm cen} \rangle = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\log_{10} M_{\rm halo} - \log_{10} M_{\rm min}}{\sigma_{\log_{10} M}}\right)\right), \tag{2.33}$$

$$\langle N_{\rm sat} \rangle = \langle N_{\rm cen} \rangle \times \left(\frac{M_{\rm halo} - M_0}{M_1}\right)^{\alpha}, \tag{2.34}$$

$$\langle N_{\rm total} \rangle = \langle N_{\rm central} \rangle + \langle N_{\rm sat} \rangle = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\log_{10} M_{\rm halo} - \log_{10} M_{\rm min}}{\sigma_{\log_{10} M}}\right)\right)\left(1 + \left(\frac{M_{\rm halo} - M_0}{M_1}\right)^{\alpha}\right), \tag{2.35}$$

[2] With the high-redshift correction of Behroozi et al. (2013) in Chapter 5.
[3] https://github.com/steven-murray/halomod

and for constructing the central-satellite term

$$\langle N_{\rm cen} N_{\rm sat} \rangle = \frac{1}{2}\left(1 + {\rm erf}\left(\frac{\log_{10} M_{\rm halo} - \log_{10} M_{\rm min}}{\sigma_{\log_{10} M}}\right)\right), \tag{2.36}$$

and the satellite-satellite term

$$\langle N_{\rm sat}(N_{\rm sat} - 1)\rangle = \langle N_{\rm sat}\rangle^2, \tag{2.37}$$

given that it is Poissonian.

The other commonly used three parameter model of Zehavi et al. (2005) is the limit of setting M_0 and $\sigma_{\log_{10} M}$ to zero.

Figure 2.8 shows how the correlation function and comoving number density evolves with each Zheng et al. (2005) HOD parameter. Each parameter has the following impact:

- $M_{\rm min}$ increasing; increases 2-halo term because galaxies are in more massive halos, so the bias increases; increases the 1-halo term as $M_{\rm min}$ is now closer to M_1 and the satellite fraction becomes greater; decreases the galaxy number density because rarer haloes are needed.

- M_1 increasing; decreases the 1-halo term as the satellite fraction decreases; decreases the 2-halo term very slightly as the average halo mass is slightly decreased; decreases number density as the number of satellites decreases.

- α increasing; increases the 2-halo term as average halo mass increases; increases the 1-halo term as higher satellite fraction (but specifically the satellite-satellite term, as the extra satellites are achieved by making massive halos either have lots of satellites, or none - as opposed to by changing the threshold halo mass for satellites); increases number density as more satellites (also increases very slightly for extremely low α as this corresponds to haloes with masses less than M_1 start to have non-trivial numbers of satellites)

- $\sigma_{\log_{10} M}$ increases; number density increases (scatter of centrals into lower mass haloes outweighs scatter into higher mass haloes because of the gradient of the halo mass

function); slightly decreases 2-halo term as average halo mass decreased; slightly decreases 1-halo term as satellite fraction decreased.

- M_0 increases; minuscule decrease in 2-halo term; minuscule change in 1-halo term until M_0 becomes comparable with M_1 then dramatic decreases as the number of satellites decreases; number density decreases very slightly (number of satellites decreases, but this is a small fraction of the total number of galaxies).

2.3.6 Incorporating Stellar Mass Ranges in the Model

Angular correlation functions, and hence the derived HOD models, are highly dependent on the magnitude cut or stellar mass range of the galaxies, which is to be expected as different galaxy samples typically exist in different halos. The approach that I use to build up a self consistent picture of how galaxies of different stellar masses occupy the halos is to calculate the correlation function for all the galaxies above a certain mass threshold, for a range of thresholds. The HOD models are then expected to be consistent e.g. a sample of a higher stellar mass threshold does not predict more galaxies at a given halo mass than a lower stellar mass threshold! An alternate approach would be to calculate the correlation function for stellar mass ranges as in Coupon et al. (2015), which reduces the covariance between measurements. This, however, is better suited to fitting a global occupation model where the halo occupation is a conditional function of the stellar mass given the halo mass, because otherwise the occupation number as a function of halo mass is not straightforward when there is an upper bound of stellar mass.

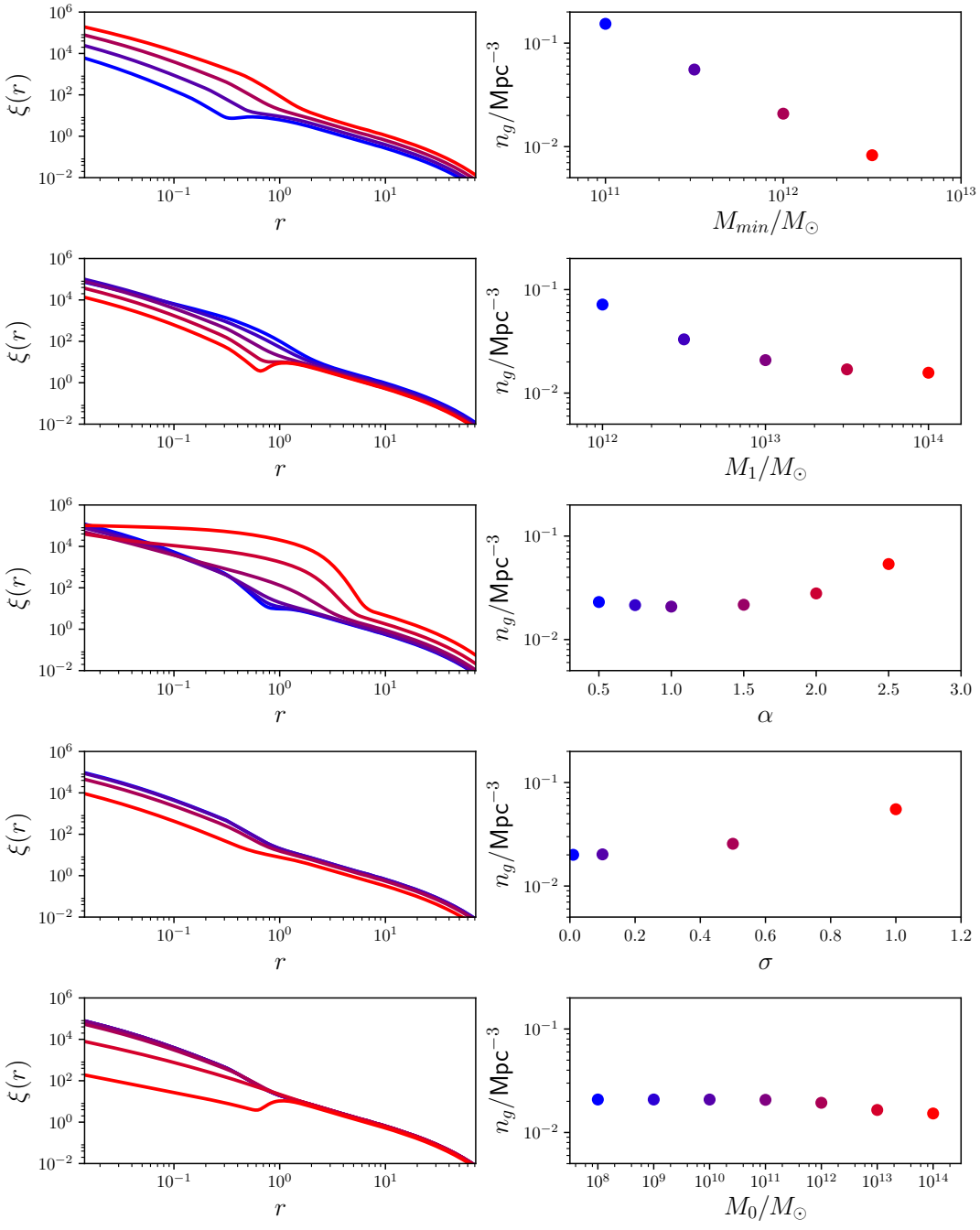


Figure 2.8: Illustration of how the spatial correlation function (left column) and galaxy comoving number density (right column) of galaxies varies with each individual Zheng et al. (2005) HOD parameter. Each model has $M_{\rm min} = 10^{12} M_\odot$, $M_1 = 10^{13} M_\odot$, $\alpha = 1$, $\sigma = 0.2$ and $M_0 = 10^8 M_\odot$, apart from the individual parameter being varied, and is for $z = 1$. Blue to red corresponds to increasing parameter value.

Chapter 3

The Galaxy-Halo Connection in the VIDEO Survey

This Chapter is based on the published paper Hatfield et al. (2016).

3.1 Introduction

As discussed in Chapters 1 and 2, contemporary galaxy evolution works in the paradigm of luminous matter (galaxies) being biased tracers of the underlying dark matter distribution. The growth of cold dark matter (CDM) perturbations is relatively simple to model and understand, both analytically (Press & Schechter, 1974; Sheth & Tormen, 1999) and in N-body simulations (Warren et al., 2006) as it is thought to be pressure and interaction free, see Chapter 1. However we cannot observe the dark matter directly; we can only observe the luminous matter following the underlying dark matter distribution in a biased, complex way. Large galaxy surveys allow us to probe this behaviour in a statistical manner, giving insight to the physical processes at play. Recent wide-field surveys have surveyed the semi-local Universe spectroscopically in great detail e.g. the 2-degree-Field Galaxy Redshift Survey (2dFGRS, Peacock et al., 2001), Sloan Digital Sky Survey (SDSS, Zehavi et al., 2011) and the Galaxy And Mass Assembly (GAMA, Driver et al., 2011) survey on the kilo-square degree scale, the VIMOS VLT Deep Survey (VVDS, Le Fèvre et al., 2013) and the VIMOS Ultra-Deep Survey (VUDS, Le Fèvre et al., 2015) on degree scales. Similarly,

surveys like the United Kingdom Infrared Deep Sky Survey Ultra Deep Survey (UKIDSS-UDS, Hartley et al., 2013) and now UltraVISTA (McCracken et al., 2012), have probed photometrically very deeply on $\sim 1\text{deg}^2$ scales. The Visible and Infrared Survey Telescope for Astronomy (VISTA) Deep Extragalactic Observations (VIDEO) survey (Jarvis et al., 2013) sits between these two scales of interest as the current leading survey for studying the $z > 0.5$ Universe over large scales. It is particularly well suited to investigating many contemporary problems in forming a good all-encompassing model of galaxy evolution. Although modern observational techniques have led to substantial improvements in our understanding of the nature of galaxies and their evolution over cosmic time (e.g. see Mo, van den Bosch and White, 2010 for a review), there remain many problems in explaining the rich menagerie of galaxies we see in the Universe today. Galaxies come in a range of masses spanning several decades (e.g. Tomczak et al. 2014), exhibit a range of morphologies (e.g. Hubble, 1926, Roberts & Haynes, 1994, Willett et al. 2013), and can have vastly different star-formation (SF) rates (e.g. Rieke et al., 1980, Bergvall et al. 2016). Some exhibit active galactic nucleus (AGN) activity - powerful energetic bursts from accretion onto supermassive black holes, that are thought to impact on the life of the whole galaxy via feedback processes (e.g. Croton et al., 2006, Fabian 2012). A good model of galaxy evolution must take all these wide ranging phenomena into account (e.g. most semi-analytic and hydrodynamic simulations now incorporate such activity to truncate star formation in massive galaxies, for example Dubois et al., 2014) to explain the observations.

VIDEO is particularly well suited to investigating, explaining and constraining many of these problems, as its balance of depth and sky area allows wide scale effects to be probed to earlier times:

- It has a multitude of multi-band data for both better constraints on redshift as well as extra information like stellar mass and star formation rate of the galaxies, see Johnston et al. (2015)
- Its depth and high quality photometric redshifts permit the study of galaxies on large scales at $z \sim 1 - 3$, the peak of star formation in the Universe
- Its balance of depth and sky area makes it possible to determine galaxy behaviour on

both sides of the 'knee' of the stellar mass function at these crucial redshifts

- It has the area coverage and resolution to simultaneously probe the two length-scale regimes of linear and non-linear distributions
- It has three separate fields to measure cosmic variance

Access to these large-scale effects is crucial for understanding the environment of a galaxy population, which can play an important role in its evolution. Key processes in galaxy evolution are often classified into 'nature' and 'nurture' effects, e.g. internal processes such as cooling and feedback versus interactions with other galaxies and the local environment - often a variety of processes are needed to explain observations such as the morphology-density relation (elliptical galaxies are preferentially found in high-density environments and spiral galaxies in the field; Dressler, 1980). A key question is the role of environment and halo mass on quenching (e.g. Peng et al., 2010), and how important, or not, processes like strangulation (tidal effects from the gravitational potential allowing the gas in the satellite to leave, Peng et al., 2015), ram pressure stripping (removal of gas by 'winds' in the hot intra-cluster medium, see Gunn & Gott, 1972 for an early description, Abadi et al., 1999 for a key simulation of the process and Steinhauser et al., 2016 for a more recent discussion) and harassment (flybys from other galaxies, Farouki & Shapiro, 1981 and Moore et al., 1998) are. It is also becoming apparent that the larger scale environment, distances well beyond the virial radius of the halo, can have local effects on individual galaxies and lead to large scale correlations, now known as galactic conformity (e.g. Weinmann et al. 2006, Kauffmann et al. 2013, Hearin et al. 2016a, Knobel et al. 2015).

One key probe of the galaxy-dark matter connection is the two-point correlation function, see Peebles (1980) and Chapter 2. This is commonly interpreted via the phenomenological model of the halo occupation distribution (HOD, see Chapter 2, Cooray & Sheth, 2002; Zehavi et al., 2005). Typically the galaxy content of a halo is stipulated as some function of the halo mass. Then assuming a halo bias model and halo profile, the correlation function can be predicted, and compared to observations (Zheng et al., 2005). Derived parameters from the HOD (minimum mass for galaxy collapse, bias, typical halo mass etc) can then typically be linked to models of galaxy formation and evolution, or compared with

results from simulations (e.g. Wang et al., 2006). Other probes of the galaxy-DM connection include galaxy-galaxy lensing, which contains information about the host halo profile and can be combined with clustering measurements to great effect (e.g. Coupon et al., 2015), and comparison with group catalogues (e.g Yang et al., 2005). In this Chapter I analyse the clustering relations between different galaxy samples to draw out HOD parameters to investigate the galaxy-halo connection to high-z and moderate stellar mass.

The only other survey currently able to probe to similar stellar masses and redshifts on degree scales is COSMOS and its near infrared counterpart UltraVISTA, another public ESO VISTA survey, see McCracken et al. (2012). McCracken et al. (2015) perform a clustering analysis in the survey, fitting HOD models, and studying the stellar mass to halo mass ratio. UltraVISTA and the sub-field of VIDEO that I use here probe similar parts of parameter space, giving VIDEO an important role in validating this science on a different field, but in future data releases the surveys will diverge, VIDEO probing wider, and UltraVISTA deeper. Validating clustering measurements on independent fields has particular importance in this instance as the COSMOS field (in which UltraVISTA is carried out) is reported in the literature to have an overabundance of rich structure, and to in general be unrepresentative of similar volumes at the same redshift (e.g Meneux et al. 2009, who report a $2-3\sigma$ anomaly by comparison with mock skies). McCracken et al. (2015) explore this complication, speculating that the quasar wall a few degrees away from the field, reported in Clowes et al. (2013), could give rise to this over-density. They compare clustering measurements in COSMOS with WIRDS data (Bielby et al. 2014) over four fields finding agreement on larger scales, but dramatically increased clustering power at small scales in COSMOS at $1 < z < 1.5$. Not only does this illustrate the importance of having a separate field to confirm these results at these key redshifts over the key epoch when both AGN and SF activity were at their peak, but it also shows that cosmic variance is still a significant factor at these angular scales and that eventually the multiple independent fields of VIDEO are needed. There is also valuable information to be gained by comparing photometry-based results with spectroscopic surveys that have covered the same fields (e.g. VVDS, Abbas et al. 2010, and VUDS, Durkalec et al. 2015a,b). Spectroscopic surveys have much more accurate redshifts, and can hence get more accurate measurements of clustering,

as well as probing effects not present in angular information, in particular redshift space distortions. Conversely, like-for-like spectroscopic surveys typically probe smaller numbers of sources (in a biased manner depending on the selection of the survey sources), ordinarily not probing as deep as an otherwise similar photometric survey. Exploiting the ability of spectroscopic surveys to probe different parts of clustering parameter space in different ways is beneficial for a comprehensive understanding of the galaxy-halo connection and the role of environmental effects at a given epoch.

This Chapter is organised as follows: first I describe my sample selection from VIDEO (section 3.2), and discuss how I measure the correlation function (section 3.3). I then discuss the halo occupation model that I use, derived parameters and fitting process in section 3.4. I then measure the correlation function for a series of sub-samples split by stellar mass, and fit HOD models to these observations. Finally I discuss how derived parameters from the HOD vary with stellar mass and redshift, compare to other studies, and discuss how my measurements will be extended with the full VIDEO survey (section 3.5).

3.2 Observations

In this section I describe the optical and near infrared data used to select the galaxies in my sample, and provide information on the photometric redshift and stellar mass estimates that underpin the analysis.

3.2.1 VIDEO and CFHTLS

The VIDEO Survey (Jarvis et al., 2013) is one of the 6 public surveys carried out by the VISTA telescope facility in Chile. It covers three fields in the southern hemisphere, each carefully chosen for availability of multiband data, to total 12 deg^2 when complete. The 5σ depths of VIDEO originally planned, and observed to in the XMM3 field, in the five bands are $Z = 25.7$, $Y = 24.5$, $J = 24.4$, $H = 24.1$ and $K_s = 23.8$ for a 2" diameter aperture. I note however that the observing plan is now to observe to $Y = 25.5$ at the expense of Z due to the inclusion of the fields in the Dark Energy Survey, (DES), see Banerji et al. (2014).

In this study, I use the VIDEO data set released in 2014 combined with data from the

T0006 release of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) D1 tile (Ilbert et al., 2006; Gwyn, 2012), which provides photometry with 5σ depths of $u^* = 27.4$, $g' = 27.9$, $r' = 27.6$, $i' = 27.4$ $z' = 26.1$ over 1 deg^2 of the VIDEO XMM3 tile (which will be joined by two other adjacent tiles). Note that the i' filter used for CHFTLS is different to the SDSS i' filter, and that this data was collected with the *first* MegaCam i' filter (during the survey the filter had to be replaced by one with a slightly different response). This data set (and the parameterisations discussed in section 3.2.2) has already been used in many extragalactic studies to data (e.g. White et al., 2015; Johnston et al., 2015). The infrared VIDEO data for the other tiles is now available. However, the publicly available optical data over these fields (CHFTLS Wide-1 and the public DES data) are shallower than D1, which would not allow extension to as high redshifts. Future work will extend the analysis in this paper to the wider areas (see Chapter 7).

3.2.2 LePHARE and SExtractor

The sources and photometry from the images are extracted using SExtractor, (Bertin & Arnouts, 1996) source extraction software, with 2" apertures. See Jarvis et al. (2013) for more details.

The photometric redshifts are calculated using LePHARE (Arnouts et al., 1999; Ilbert et al., 2006), which fits spectral energy distribution (SED) templates (constructed using SPS models, see section 1.3.1.4) to the photometry (Jarvis et al., 2013). LePHARE generates a redshift probability density function, stellar masses, star formation rates, CLASS_STAR (probability of being a star based on compactness) along with many other parameters. Further information on detection images used, detection thresholds and the construction of the SED templates is given in Jarvis et al. (2013). The VIDEO LePHARE photometric redshifts were compared and validated against spectroscopic redshifts from VVDS (Le Fèvre et al., 2013), see figure 11 in Jarvis et al. (2013). Jarvis et al. (2013) found that the VIDEO+CHFTLS ten-band photometric redshifts had a normalised median absolute deviation of $\frac{|z_{spec}-z_{phot}|}{1+z_{spec}} \sim 0.023$ and a catastrophic outlier fraction ($\frac{|z_{spec}-z_{phot}|}{1+z_{spec}} > 0.15$) of ~ 3.8 percent. LePHARE has been used in recent studies of the stellar mass function as a function of redshift (Ilbert et al., 2015) and is widely used throughout the community.

However, I note Laigle et al. in prep. (see also Chapter 6) shows that there may be systematic offsets in stellar masses due to incomplete understanding of star formation histories. See Abdalla et al. (2011) for a discussion of the performance of LePHARE and similar photometric redshift codes.

3.2.3 Final Sample

SExtractor identifies 481,685 sources in the 1 deg^2 field with detections in at least one band. I applied a simple mask to the data in order to cut out areas dominated by foreground stars and any dead pixels. The mask was also applied to the random catalogues used in the calculation of the correlation function (see section 2.1.2).

Uncertainty in LePHARE parameterisations (photometric redshift estimation etc.) increases at fainter magnitudes, both because the relative error on fluxes is larger for faint objects, and because objects start to be only detected in a few bands. I use a K_s-band cut to remove all galaxies with K_s >23.5. VIDEO has a 90 percent completeness at this depth (Jarvis et al., 2013).

For removing stars from the sample, SExtractor produces parameter CLASS_STAR as an indicator of the probability that a given object is a star, based on whether it appears point-like, but this has been shown not to perform well up to the magnitudes I use (McAlpine et al., 2012, White et al., 2015). To eliminate stars from the sample I defined a stellar locus as in Jarvis et al. (2013), following the approach of Baldry et al. (2010),

$$f_{\rm locus}(x) = \begin{cases} -0.58 & x < 0.4 \\ -0.58 + 0.82x - 0.21x^2 & 0.4 < x < 1.9 \\ -0.08 & 1.9 < x \end{cases} . \quad (3.1)$$

I then remove sources with:

$$J - K_s < 0.12 + f_{\rm locus}(g - i). \quad (3.2)$$

McAlpine et al. estimate this cut leaves stars contributing less than 5% of the sample. The final galaxy sample comprises 97,052 sources after masking, removing stars and making

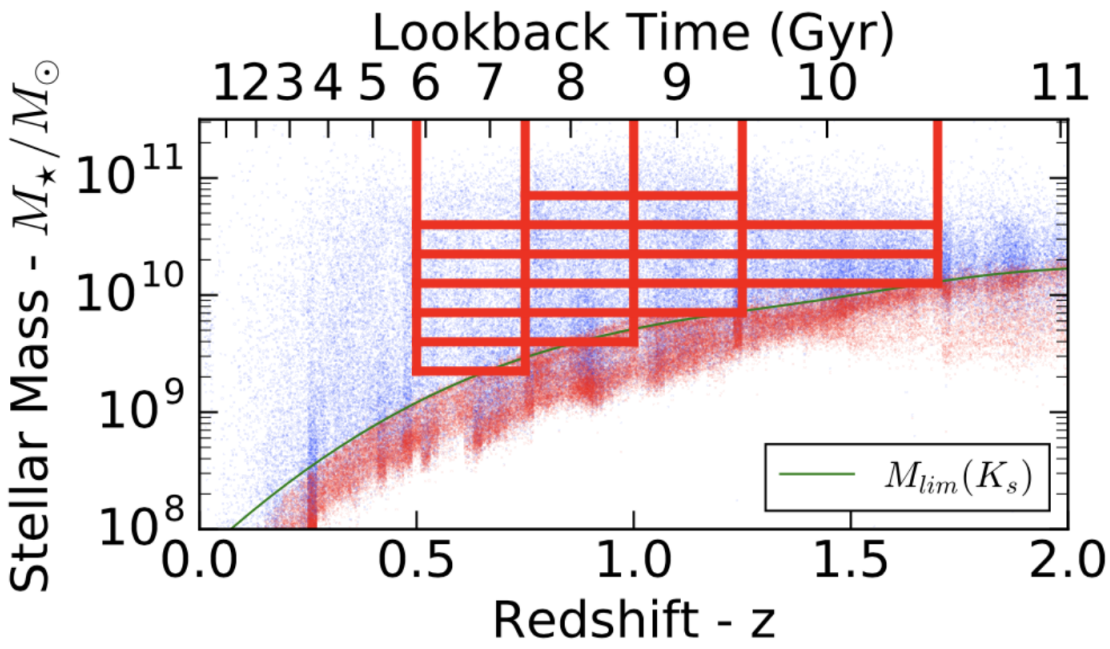


Figure 3.1: The mass and redshift of galaxies, after application of the magnitude cut, star exclusion and mask, are shown here in blue. The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of $K_s < 23.5$, and the green curve the implied 90% stellar mass completeness limit, following the approach of Johnston et al. (2015). The red boxes illustrate the redshift and stellar mass selected sub-samples that I consider in subsequent sections.

a K_s <23.5 cut.

3.3 The Two-Point Correlation Function

As discussed in Chapter 2, a range of statistical tools are used to study the interactions between galaxies and to characterise clustering. In this Chapter I focus on the angular two-point correlation function, a measure of how much more likely two galaxies are to be at a given angular separation than random.

I calculate the angular correlation function as discussed in Chapter 2. I use the Landy & Szalay (1993) estimator, weighting galaxies according to their redshift probability distribution functions as per Arnouts et al. (2002), estimate the correlation function continuously as per section 2.1.2.3, and include the integral constraint in the model as per Beutler et al. (2011), as opposed to correcting our data points - complete details of this process are given

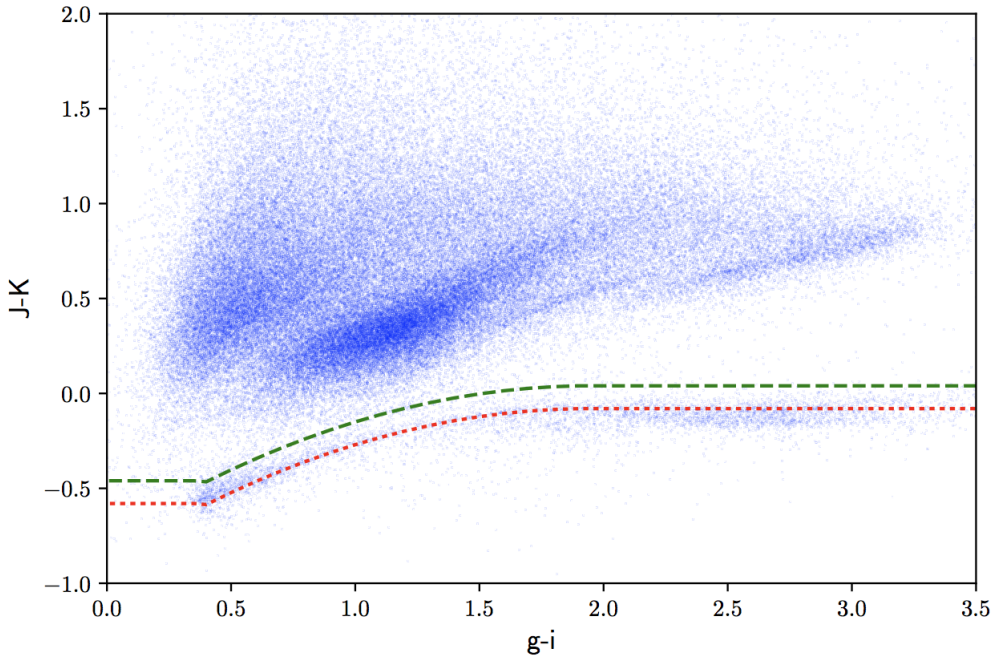


Figure 3.2: Colour-colour diagram of all the K< 23.5 sources in my sample, with stellar locus (red dotted line), and stellar cut (green dashed line) shown.

in Chapter 2.

3.4 Halo Occupation Distribution Modelling

In this Chapter I model the clustering measurements and number counts using a HOD model, as per Chapter 2.

3.4.1 The Model

I use the 5 parameter model of Zheng et al. (2005), assuming a Navaro-Frenk-White profile (Navarro, Frenk, & White, 1997) and a Tinker et al. (2010) bias and HMF model, see Chapter 2. I use the HALOMOD python package[1] to calculate model correlation functions and derived parameters.

I also calculate for completeness r_0, the comoving separation at which the spatial correlation function (for the best fit parameters) is unity. This is useful as it operates as a one-dimensional measurement of clustering (as opposed to HOD parameters, which cannot

[1] https://github.com/steven-murray/halomod

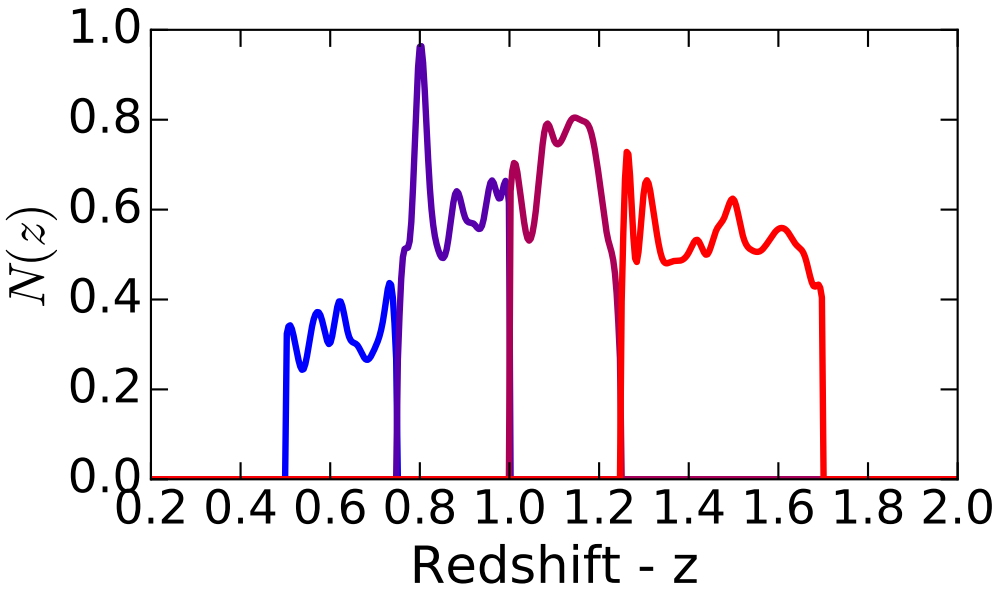


Figure 3.3: The redshift distributions used in the analysis for each redshift bin (arbitrary normalisation).

be summarised with one number). It also allows comparison with studies that study correlation functions with a power law, which normally model the spatial correlation function as $\xi(r) = (r/r_0)^{-\gamma}$. Note r_0 in the context of this Chapter is a derived parameter of the HOD model; it does not come from a power law fit to the correlation function.

3.4.2 Projecting from $3D$ and Choice of $N(z)$

Projecting the spatial correlation function to angular space requires input of the redshift distribution of the galaxies in the sample, $N(z)$,. As discussed in Chapter 2, I use a system of weights when calculating the correlation function. Therefore I use the sum of the pdfs just in the redshift range considered, which leads to sharp cutoffs, see fig 3.3.

I note a sharp peak at $z \sim 0.8$, which could indicate the presence of a large structure at this redshift.

3.4.3 MCMC Fitting Process

I use EMCEE[2] (Foreman-Mackey et al., 2013) to provide a Markov Chain Monte Carlo (MCMC) sampling of the parameter space to fit the correlation functions. EMCEE uses a variation of the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), see Foreman-Mackey et al. (2013) for details. I use a uniform prior over $0.5 < \alpha < 2.5$, $0 < \sigma < 0.6$, $10 < \log_{10}(M_{\rm min}/M_\odot) < 15$, $\log_{10}(M_{\rm min}/M_\odot) < \log_{10}(M_1/M_\odot) < 17$ and $8 < \log_{10}(M_0/M_\odot) < \log_{10}(M_1/M_\odot)$ (uniform in log space for mass). I used 20 walkers with 1000 steps, which have starting positions drawn uniformly from the prior. To test for the convergence of the MCMC sampling, for a subset of the correlation functions I repeated the analysis with ten times the number of steps. Using longer chains did not change my parameter estimation, increasing my confidence that my MCMC runs had reached convergence[3].

I calculated the likelihood using χ^2 from both the correlation function and the galaxy abundance,

$$\chi^2 = \frac{[N_{\rm gal}^{\rm obs} - N_{\rm gal}^{\rm model}]^2}{\sigma_n^2} + \sum_i \frac{[\omega^{\rm obs}(\theta_i) - \omega^{\rm model}(\theta_i)]^2}{\sigma_{w_i}^2}, \tag{3.3}$$

where $N_{\rm gal}^{\rm obs}$ is the observed number of galaxies in the sample, $N_{\rm gal}^{\rm model}$ is the predicted number of galaxies in that redshift range for a given model, σ_n is the error on the number counts including both Poisson noise and cosmic variance, θ_i are the angular scales that are fit, $\omega^{\rm obs}$ is the observed angular correlation function, $\omega^{\rm model}$ is the angular correlation function of a given model, and σ_{w_i} is the error on the measured correlation function from the bootstrapping[4].

As I estimate $\omega(\theta)$ as a continuous function, covariance between measurements is less straightforward. I work around this by fitting to points equally separated in log-space between 0.001° and 0.1°, with the separation chosen to be greater than the smoothing scale of the non-parametric estimation to minimise covariance between points. I calculate the

[2] http://dan.iel.fm/emcee/current/

[3] A similar approach to testing for convergence is used in subsequent chapters.

[4] Covariance between measurements of $\omega^{\rm obs}$ at different θ_i is neglected as the covariances are much smaller than the σ_{w_i}. The covariances become more significant in Chapter 5, where they are accounted for in the likelihood.

error on the number counts (which must include both Poisson noise and cosmic variance) as in Trenti & Stiavelli (2008).

Finding $N_{\rm gal}^{\rm obs}$ is complicated by the mask used to remove defects in the field (discussed section 3.2.3), as well as the fact that each galaxy is effectively in multiple redshift bins. I account for this by instead of counting the galaxies, counting the weights, and then rescaling by the amount of field lost by the mask:

$$N_{\rm gal}^{\rm obs} = \frac{\sum\limits_{i} W_i}{1-A}, \tag{3.4}$$

where A is the fraction of the field covered by the mask (0.03 in this case) and W_i are the weights from equation 2.20.

3.5 Results

3.5.1 Redshift and Stellar Mass Selection

I divided the data into four redshift bins ($0.50 < z < 0.75$, $0.75 < z < 1.00$, $1.00 < z < 1.25$ and $1.25 < z < 1.70$) and seven mass bins ($10^{9.35} M_\odot < M_\star$, $10^{9.6} M_\odot < M_\star$, $10^{9.85} M_\odot < M_\star$, $10^{10.1} M_\odot < M_\star$, $10^{10.35} M_\odot < M_\star$, $10^{10.6} M_\odot < M_\star$ and $10^{10.85} M_\odot < M_\star$), see figure 3.1. For each bin, I calculated the angular correlation function as described in section 2.1.2. Figure 3.4 shows the measurements. In each bin, near power law behaviour is visible, with some bins suggestive of the kink associated with the transition from the 1-halo to the 2-halo term, although the impact of the integral constraint at large angular scales means this transition is unlikely to be evident until there is access to the larger angular scales present in the full VIDEO survey. The clear trend of clustering increasing with stellar mass threshold at all scales is visible in all redshift bins. After the correlation function was calculated, I fit a HOD model for each subsample, as described in section 3.4.

In general, good fits to the data were obtained, see table 3.1. Figure 3.5 shows the correlation functions and corresponding best fit models. The fits had χ^2/d.o.f. values (showing in table 1) between 0.228 and 1.66, with the exception of the $0.75 < z < 1$, $10^{10.85} M_\odot < M_\star$ bin, which had χ^2/d.o.f. = 4.33. This suggests the data was well described by the HOD

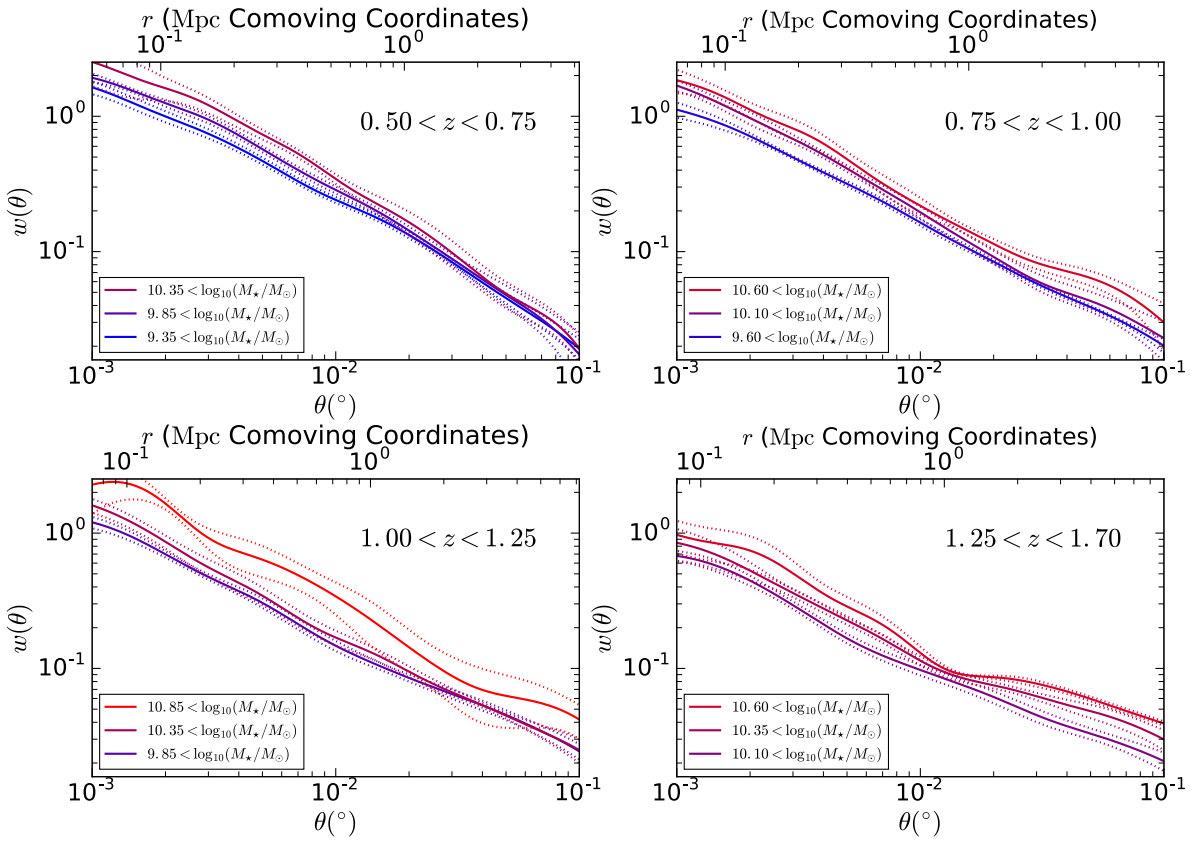


Figure 3.4: The angular correlation function for different redshift ranges and masses denoted in each panel. The lower x-axis denotes angular scale, the upper x-axis the corresponding projected comoving distance, and the y-axis the correlation function. The fainter, dashed, upper and lower bands represent the error bars on the measurements, discussed in section 2.1.2.3. For the clarity of the plot, I only show alternate stellar mass samples in the first three redshift bin subplots.

model in all cases apart from one bin. Consistent with these χ^2/d.o.f. results, the outlier bin in question is seen in figure 3.5 to have a correlation function that doesn't fit into the pattern of measurements in the other mass ranges in the same redshift bin - the massive galaxies at this redshift appear to have an unusual spatial distribution (conceivably associated with the sharp peak in the redshift distribution at $z \sim 0.8$). Since there was still enough data to make a good measurement of the clustering, and it was possible to make a moderate fit, I still include the results from this bin in subsequent plots, but note the results for this bin might be subject to some unknown systematic effect.

Table 3.1: Parameters from the angular correlation function evaluation and HOD fitting, with corresponding χ^2/d.o.f. values for the fits.

Stellar Mass Threshold	N_g	M_{min}	M_1	α	σ	M_0	b	f_s	r_0	χ^2/d.o.f.
$0.50 < z < 0.75$	$z_{med} = 0.62$									
9.35	6535	$11.7^{+0.063}_{-0.075}$	$12.9^{+0.13}_{-0.21}$	$0.948^{+0.1}_{-0.16}$	$0.5^{+0.078}_{-0.2}$	$11.8^{+0.46}_{-1.3}$	$1.17^{+0.018}_{-0.012}$	$0.213^{+0.044}_{-0.019}$	$6.21^{+0.12}_{-0.085}$	1.33
9.6	5061	$11.8^{+0.089}_{-0.089}$	$12.8^{+0.22}_{-0.15}$	$0.82^{+0.15}_{-0.14}$	$0.412^{+0.14}_{-0.23}$	$12.3^{+0.19}_{-0.34}$	$1.22^{+0.024}_{-0.017}$	$0.196^{+0.014}_{-0.016}$	$6.56^{+0.17}_{-0.12}$	0.661
9.85	3877	$12.0^{+0.083}_{-0.071}$	$13.1^{+0.16}_{-0.22}$	$0.948^{+0.14}_{-0.19}$	$0.455^{+0.11}_{-0.17}$	$12.1^{+0.41}_{-1.4}$	$1.24^{+0.022}_{-0.016}$	$0.191^{+0.033}_{-0.018}$	$6.71^{+0.14}_{-0.11}$	0.918
10.1	2847	$12.1^{+0.053}_{-0.059}$	$13.1^{+0.15}_{-0.23}$	$0.887^{+0.16}_{-0.21}$	$0.539^{+0.047}_{-0.13}$	$12.2^{+0.4}_{-1.5}$	$1.26^{+0.016}_{-0.013}$	$0.193^{+0.061}_{-0.021}$	$6.92^{+0.12}_{-0.098}$	1.66
10.35	1884	$12.3^{+0.061}_{-0.084}$	$13.4^{+0.11}_{-0.2}$	$0.996^{+0.11}_{-0.3}$	$0.507^{+0.072}_{-0.18}$	$11.9^{+0.76}_{-2.2}$	$1.33^{+0.024}_{-0.019}$	$0.175^{+0.073}_{-0.026}$	$7.38^{+0.19}_{-0.14}$	1.27
10.6	1022	$12.5^{+0.088}_{-0.09}$	$13.7^{+0.087}_{-0.19}$	$1.23^{+0.14}_{-0.35}$	$0.418^{+0.14}_{-0.23}$	$12.0^{+0.95}_{-2.5}$	$1.47^{+0.036}_{-0.034}$	$0.129^{+0.026}_{-0.018}$	$8.52^{+0.27}_{-0.26}$	0.839
$0.75 < z < 1.00$	$z_{med} = 0.88$									
9.6	9791	$11.7^{+0.052}_{-0.08}$	$12.9^{+0.17}_{-0.24}$	$0.942^{+0.16}_{-0.19}$	$0.516^{+0.065}_{-0.21}$	$11.9^{+0.42}_{-1.4}$	$1.24^{+0.022}_{-0.013}$	$0.155^{+0.028}_{-0.015}$	$5.68^{+0.13}_{-0.076}$	0.862
9.85	7365	$11.8^{+0.049}_{-0.058}$	$12.9^{+0.27}_{-0.22}$	$0.796^{+0.27}_{-0.16}$	$0.529^{+0.055}_{-0.13}$	$12.3^{+0.21}_{-0.23}$	$1.27^{+0.017}_{-0.023}$	$0.149^{+0.015}_{-0.014}$	$5.89^{+0.7}_{-0.073}$	0.896
10.1	5453	$12.0^{+0.069}_{-0.069}$	$13.0^{+0.22}_{-0.22}$	$0.806^{+0.24}_{-0.19}$	$0.523^{+0.059}_{-0.18}$	$12.3^{+0.26}_{-0.75}$	$1.32^{+0.023}_{-0.017}$	$0.144^{+0.017}_{-0.013}$	$6.17^{+0.14}_{-0.1}$	1.2
10.35	3824	$12.1^{+0.061}_{-0.049}$	$13.4^{+0.11}_{-0.19}$	$0.92^{+0.19}_{-0.29}$	$0.53^{+0.052}_{-0.14}$	$12.1^{+0.43}_{-2.1}$	$1.35^{+0.021}_{-0.015}$	$0.119^{+0.043}_{-0.013}$	$6.35^{+0.13}_{-0.093}$	1.43
10.6	2330	$12.3^{+0.074}_{-0.082}$	$13.7^{+0.14}_{-0.16}$	$0.965^{+0.21}_{-0.32}$	$0.426^{+0.12}_{-0.18}$	$12.3^{+0.42}_{-1.9}$	$1.46^{+0.029}_{-0.025}$	$0.0945^{+0.026}_{-0.013}$	$7.11^{+0.19}_{-0.16}$	0.549
10.85	1023	$12.6^{+0.03}_{-0.038}$	$14.2^{+0.22}_{-0.18}$	$0.633^{+0.21}_{-0.11}$	$0.585^{+0.012}_{-0.036}$	$12.3^{+0.27}_{-1.7}$	$1.52^{+0.017}_{-0.015}$	$0.101^{+0.12}_{-0.018}$	$7.52^{+0.11}_{-0.1}$	4.33
$1.00 < z < 1.25$	$z_{med} = 1.12$									
9.85	7512	$11.8^{+0.11}_{-0.084}$	$13.2^{+0.15}_{-0.074}$	$1.21^{+0.074}_{-0.14}$	$0.339^{+0.17}_{-0.2}$	$10.4^{+1.5}_{-1.4}$	$1.46^{+0.026}_{-0.027}$	$0.13^{+0.029}_{-0.022}$	$6.09^{+0.14}_{-0.15}$	0.229
10.1	5529	$12.0^{+0.063}_{-0.052}$	$13.2^{+0.14}_{-0.24}$	$1.04^{+0.22}_{-0.29}$	$0.467^{+0.097}_{-0.2}$	$12.2^{+0.42}_{-1.1}$	$1.5^{+0.029}_{-0.021}$	$0.115^{+0.018}_{-0.016}$	$6.34^{+0.17}_{-0.11}$	0.521
10.35	3892	$12.2^{+0.071}_{-0.071}$	$13.4^{+0.062}_{-0.06}$	$1.15^{+0.082}_{-0.19}$	$0.535^{+0.049}_{-0.11}$	$11.1^{+1.1}_{-1.8}$	$1.53^{+0.021}_{-0.018}$	$0.115^{+0.037}_{-0.023}$	$6.48^{+0.12}_{-0.11}$	0.689
10.6	2412	$12.3^{+0.05}_{-0.089}$	$13.7^{+0.13}_{-0.074}$	$1.07^{+0.13}_{-0.17}$	$0.532^{+0.052}_{-0.18}$	$10.7^{+1.3}_{-1.7}$	$1.62^{+0.036}_{-0.023}$	$0.13^{+0.066}_{-0.042}$	$7.06^{+0.21}_{-0.14}$	0.937
10.85	1064	$12.6^{+0.091}_{-0.091}$	$13.9^{+0.29}_{-0.11}$	$1.25^{+0.28}_{-0.5}$	$0.378^{+0.17}_{-0.21}$	$11.8^{+1.1}_{-2.1}$	$1.93^{+0.058}_{-0.054}$	$0.068^{+0.021}_{-0.012}$	$8.95^{+0.36}_{-0.34}$	0.635
$1.25 < z < 1.70$	$z_{med} = 1.48$									
10.1	10800	$11.9^{+0.065}_{-0.036}$	$13.3^{+0.057}_{-0.12}$	$1.36^{+0.11}_{-0.32}$	$0.188^{+0.15}_{-0.13}$	$11.4^{+0.98}_{-2.0}$	$1.78^{+0.021}_{-0.027}$	$0.076^{+0.011}_{-0.0087}$	$6.47^{+0.11}_{-0.14}$	0.437
10.35	5875	$12.1^{+0.053}_{-0.033}$	$13.5^{+0.075}_{-0.16}$	$1.41^{+0.17}_{-0.46}$	$0.119^{+0.14}_{-0.085}$	$11.8^{+0.88}_{-2.4}$	$1.98^{+0.027}_{-0.026}$	$0.0575^{+0.0081}_{-0.0076}$	$7.54^{+0.14}_{-0.14}$	0.932
10.6	2542	$12.4^{+0.068}_{-0.068}$	$14.0^{+0.45}_{-0.16}$	$1.06^{+0.38}_{-0.43}$	$0.329^{+0.13}_{-0.15}$	$12.2^{+0.67}_{-2.2}$	$2.16^{+0.036}_{-0.044}$	$0.0369^{+0.016}_{-0.0076}$	$8.52^{+0.2}_{-0.24}$	1.41

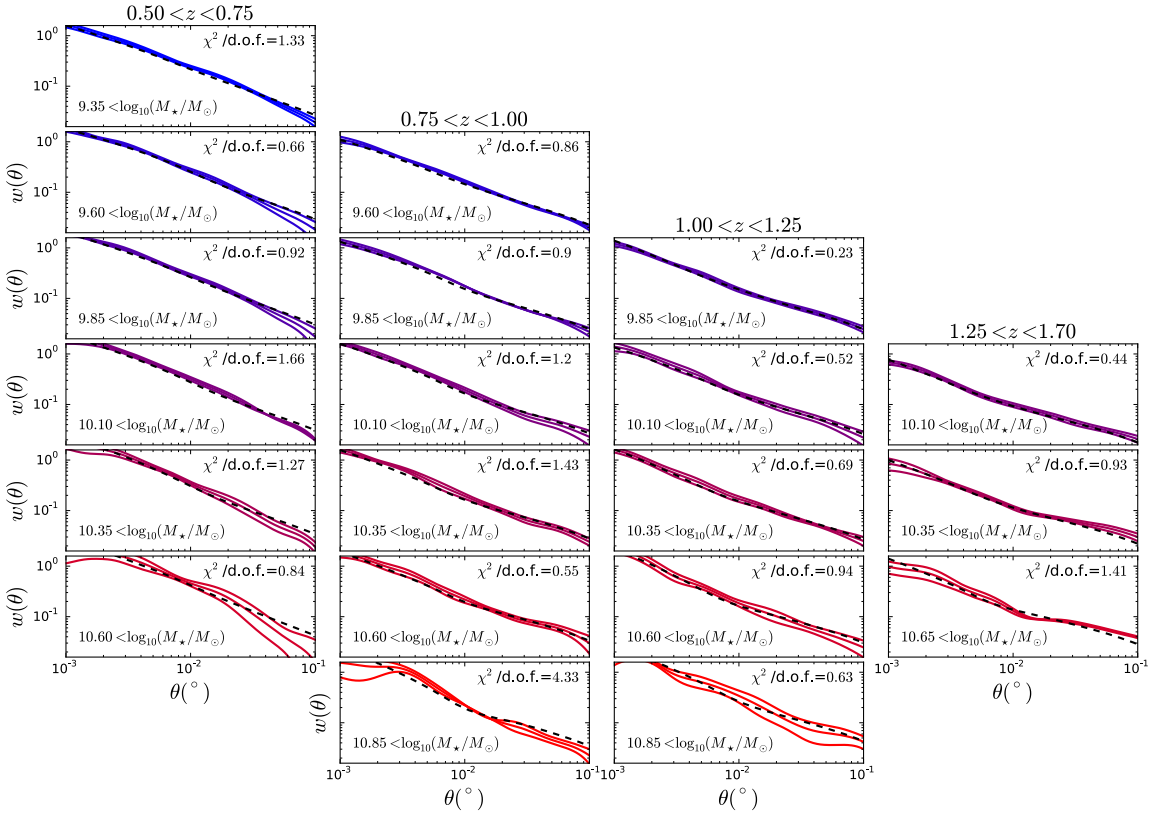


Figure 3.5: The measured correlation functions and the corresponding HOD best fits. Sub-figures in the same column have the same redshift, sub-figures in the same row have the same stellar mass range. The coloured filled lines are the data (blue to red corresponding to increasing stellar mass), and the lower and upper bands are 16th and 84th percentiles from the bootstrapping. The dashed black line is the model correlation function from the best fits. χ^2 values for each fit are shown in the upper right of each sub-figure.

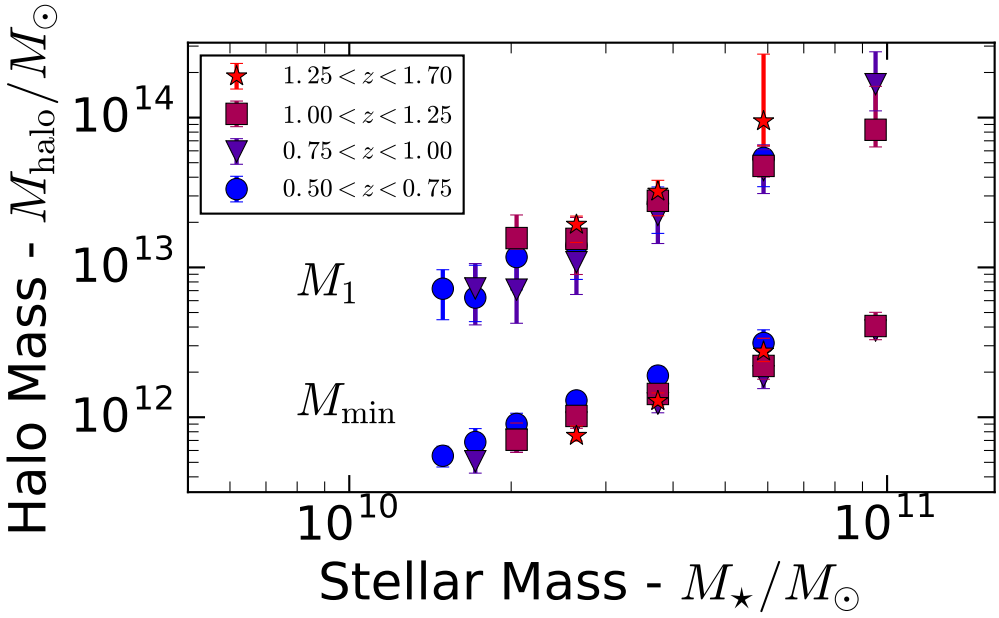


Figure 3.6: Evolution of the HOD parameter $M_{\rm min}$ and M_1 as a function of stellar mass in the four redshift bins denoted in the legend.

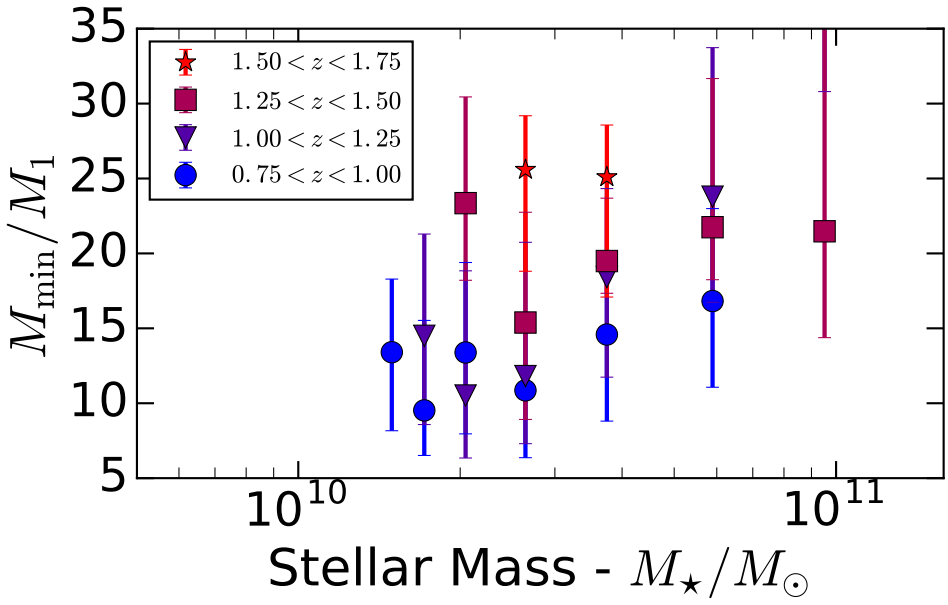


Figure 3.7: Evolution of the ratio of the HOD parameters M_1 to $M_{\rm min}$ as a function of stellar mass in the four redshift bins denoted in the legend.

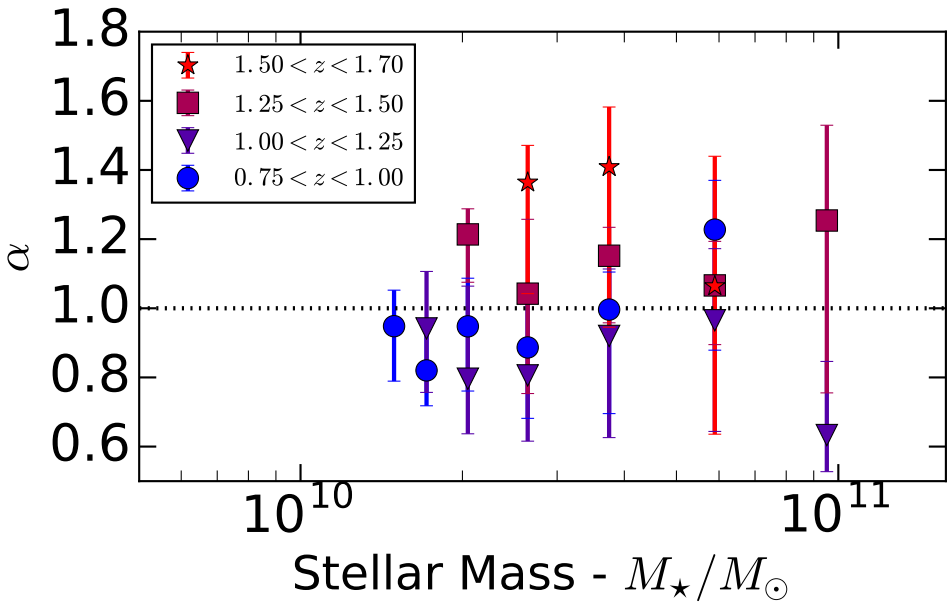


Figure 3.8: Evolution of the HOD parameter α as a function of stellar mass in the four redshift bins denoted in the legend. The dotted line represents $\alpha = 1$.

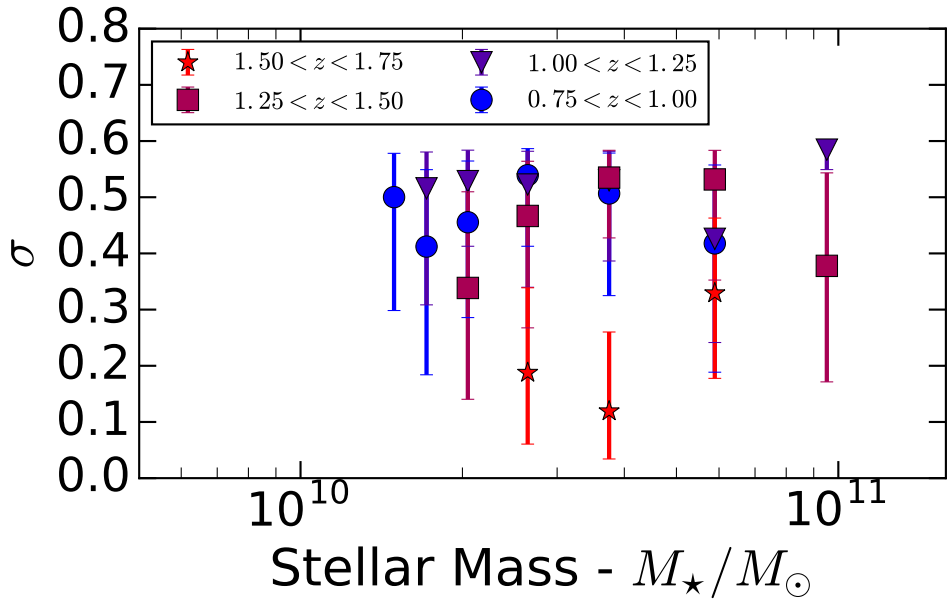


Figure 3.9: Evolution of the HOD parameter σ as a function of stellar mass in the four redshift bins denoted in the legend

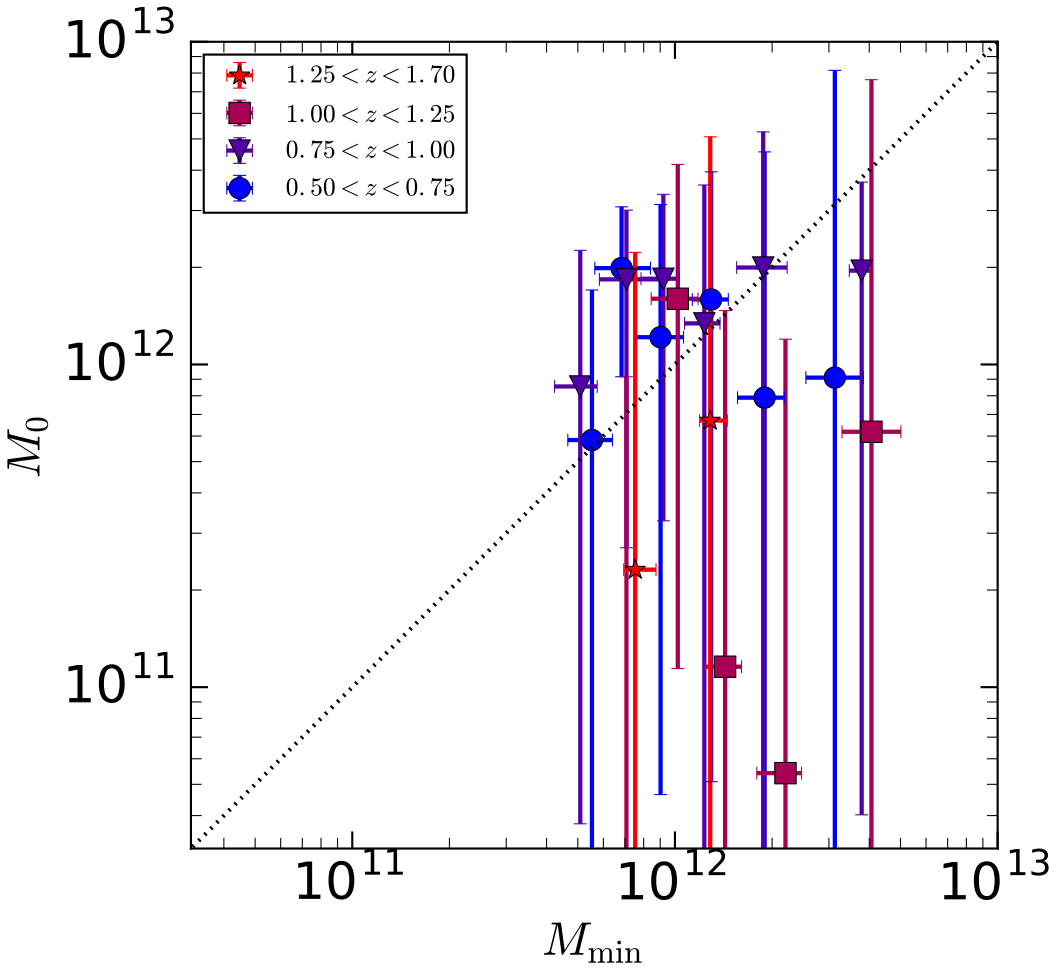


Figure 3.10: Plot comparing HOD parameters $M_{\rm min}$ and M_0 in the four redshift bins denoted in the legend. A one to one line is over-plotted as a guide.

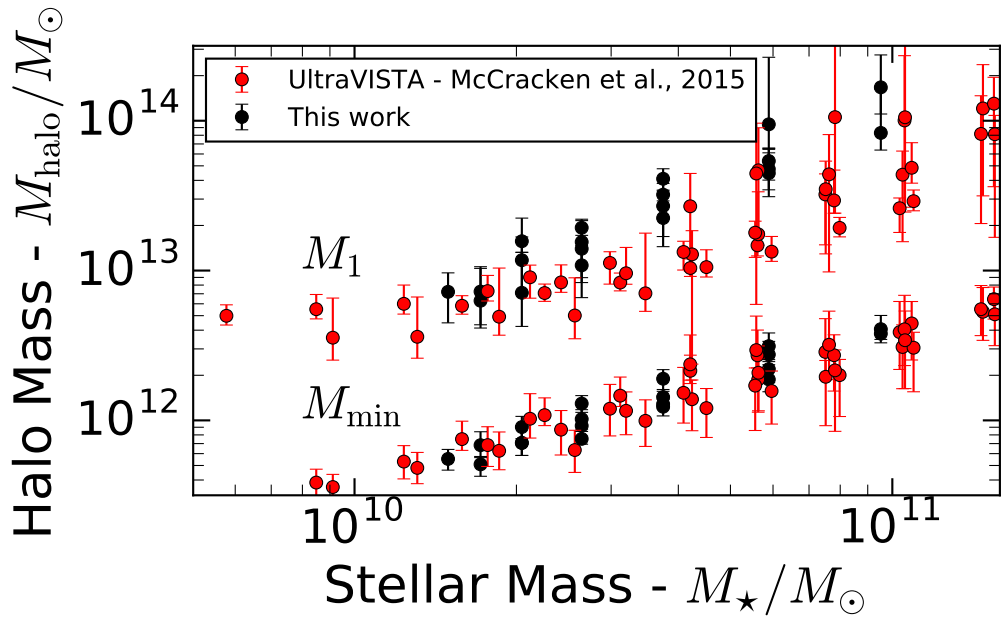


Figure 3.11: Comparison of my HOD results with similar modelling of UltraVISTA data in McCracken et al. (2015).

3.5.2 $M_{\rm min}$, M_1 and the Mass Gap

Figure 3.6 shows both the minimum halo mass for a central galaxy, $M_{\rm min}$, and the halo mass required to have satellites, M_1, growing as approximate power laws with median stellar mass, with little to no evidence of redshift evolution; this shows the well known result that more massive galaxies reside in more massive halos. M_1 remains slightly more than an order of magnitude more massive than $M_{\rm min}$ over all stellar masses. I do not detect any upturn in M_1 at stellar masses $> 10^{10.5} M_\odot$ seen in McCracken et al. (2015), although this is perhaps not surprising as I do not reach to as high masses as UltraVISTA (due to the slightly smaller area used in this paper), and based on their results I would only expect to see the upturn in the highest mass range.

Figure 3.7 shows $M_1/M_{\rm min}$, the 'mass gap' between forming the first galaxy and forming the second - colloquially how 'hard' it is for satellites to form. A high $M_1/M_{\rm min}$ means it is very hard to form satellites, and galaxies entering the halo preferentially fall onto the central galaxy rather than remaining as satellites, and vice versa. In the stellar mass ranges considered, the ratio appears to be constant with moderate scatter, potentially increasing

at higher stellar masses. Again, little redshift evolution is evident, as expected from figure 3.6. The data does not allow me to probe to low enough masses to see if there is an upturn at the lowest masses as in McCracken et al. (2015). I do find however that this ratio ranges from around 10-20 at these stellar masses (albeit with large error bars), whereas McCracken et al. (2015) finds the ratio to be around 5-10 for the same stellar masses. McCracken et al. suggest that their results help explain why different literature results measure different ratios - that different surveys are biased towards differing stellar masses, and thus correctly obtain different results. However this does not explain the discrepancy between our two results, as I explicitly control for stellar mass (although other surveys do report results not dissimilar to those found here e.g. Zehavi et al., 2011). I suggest that this is likely due to a combination of cosmic variance between the fields and the fact that for both VIDEO and UltraVISTA the field sizes only allow access to a small part of the 2-halo term, in the part of angular space where it is most difficult to account for the integral constraint. Figure 3.11 shows the results from this chapter plotted alongside the UltraVISTA results, showing close agreement, apart from the high stellar mass end for M_1, where I find slightly higher values. VUDS covered both the COSMOS field (that UltraVISTA covers) and largely overlapped the D1 field studied in this paper (in the VUDS VVDS-02h field). Durkalec et al. (2015b) report, for samples with otherwise identical selection, slightly more clustering power at larger scales in the COSMOS field, so slight clustering variance between these fields is not without precedence. I anticipate the origin of the discrepancy will become more clear with the full VIDEO survey.

3.5.3 α and σ

Figure 3.8 shows α for the four redshift bins, which parametrises how the number of satellites grows with stellar mass. I find very little evolution with either redshift or stellar mass, with typical best fit values of $\sim$1, which can be interpreted as the number of subhalos growing in proportion to the halo mass, which is to be expected. Although there appears to be some weak evolution towards higher α values at high redshift, I am cautious to claim a trend for the following reasons. Firstly the trends are of order of the size of the error bars. Secondly, measurements in the same redshift bin are not independent, so trends seen for

all the stellar mass values in a given redshift bin relative to another redshift bin are not necessarily significant. Thirdly each redshift bin is a spatially separated part of the Universe, and comparison with UltraVISTA suggests the variation between redshift bins is of a similar order to that expected from cosmic variance. In addition, the five HOD parameters are not independent from each other for a given sample, so samples having outlier values of α can have corresponding outlier values of the other parameters etc.

Figure 3.9 shows σ, which parametrises how steep the step jump in halo mass is to form the first galaxy at $M_{\rm min}$, equivalently the scatter in halo mass at fixed stellar mass for central galaxies. I find no substantial redshift or stellar mass dependence, measuring a constant value of around 0.3-0.5. Note that in some samples the posterior of σ pushes close to the boundary of the prior. Both Coupon et al. (2012) and McCracken et al. (2015) report similar findings, Coupon et al. (2012) suggesting sample incompleteness due to photometric errors could lead to missing central galaxies and hence high scatter. Alternatively it could be the case that the $z = 0$ motivated 5-parameter model I use here is less appropriate at higher redshifts. Zheng et al. (2005) give an interpretation of σ in terms of the scatter in the stellar mass at fixed halo mass: if the functional form for the number of central galaxies is an error function, then at a fixed halo mass the distribution of $\log(M_{\rm gal})$ is Gaussian. For $M_\star \propto M_{\rm halo}^{\mu}$ at that halo mass, the galaxy mass scatter can be expressed by $\mu \times \sigma = \sigma_{M_{\rm gal}}$, where $\sigma_{M_{\rm gal}}$ is the scatter in stellar mass at fixed halo mass. Using stellar mass threshold for stellar mass, and $M_{\rm min}$ for halo mass (Coupon et al., 2012), I measure $\mu \approx 2$ and thus $\sigma_{M_{\rm gal}} \approx 0.8$. An alternate way of probing the halo mass to stellar mass ratio is the Tully-Fisher relation (Tully & Fisher 1977), which relates rotational velocity to galaxy luminosity for spiral galaxies. The rotational velocity can give a measure of the dynamical mass, dominated by the halo mass, and modern stellar models can convert the luminosity into a measure of stellar mass (e.g. the redshift and mass according to the VIDEO photometry). Zheng et al. (2005) found a value of $\sigma = 0.15$ in their SPH simulation, in good agreement with dispersion measurements in the Tully-Fisher relation at $z = 0$. Tiley et al. (2016) find a greatly increased scatter at $z \sim 1$ relative to $z = 0$ in the KMOS Redshift One Survey (KROSS), reporting scatter of 0.32 dex in stellar mass at fixed stellar dynamical mass for their full sample. Although direct comparisons are difficult as the correspondence

between dynamical mass and halo mass is non-direct, and the selection methods between wide-field and integral field surveys are very different, the picture of increasing scatter at higher redshifts suggested by the two methods qualitatively agree. Subhalo abundance matching (SHAM) techniques also report dispersion of the order of ~ 0.2 at low redshift e.g. $z = 0.05$ in Reddick et al. (2013), however, as discussed in that paper, the scatter is partially an underlying assumption of the technique as opposed to a measurement.

3.5.4 Insight into Substructure

Kravtsov et al. (2004) suggest M_0 can be interpreted in terms of halo substructure. Below this quantity no satellites form, leading to the satellite occupation number to drop off more sharply than a power law. Within the paradigm of satellite galaxies living in subhalos, M_0 can be viewed as the mass at which the halo is big enough to have enough substructure to have subhalos capable of hosting their own galaxies. Sub-halo abundance matching methods assume that the stellar mass to halo mass ratio is unchanged for sub-halos. I assume that here, $M_{\rm min}$ can be viewed as the typical halo mass for that stellar mass, and M_0 can be viewed as the minimum halo mass to have a sub halo of mass $M_{\rm min}$ (and hence a satellite galaxy of that stellar mass), and compare them in figure 3.10. This shows that the constraints on M_0 are in general poor, but that M_0 is typically only slightly larger than $M_{\rm min}$, suggesting that at these times halos were rich in substructure, and could have sub-halos close in size to the whole halo. Conversely, a value of M_0 significantly larger than $M_{\rm min}$ would have suggested that halos in the given epoch were poor in substructure, and that all subhalos were dramatically smaller than the mass of the whole halo. This is in contrast to measurements at $z = 0$, by which time much of this substructure is destroyed by tidal stripping, and dynamical friction has slowed their orbits until they fall into the centre (e.g. Zentner et al., 2005), and one would see a M_0 much larger than $M_{\rm min}$.

3.5.5 Derived Parameters

The measurements of the bias b is shown in figure 3.12 and there is a clear decrease in b towards low-z (as galaxies become better tracers of the underlying dark matter and more closely follow its large scale distribution). The established trend of bias increasing with

galaxy mass is also evident; corresponding to more massive galaxies preferentially forming in larger dark matter halos, which are themselves more highly biased towards denser regions of dark matter. Note that a decrease in bias with time alone is difficult to interpret. It could be from galaxy populations moving to less massive, less biased halos over time, shifting the median host halo mass to lower values, or it could be merely from the halo bias evolution. Only with a full HOD analysis can one see that the decreasing bias is predominantly from the latter, that bias-redshift trends are driven by the halos becoming better tracers of the overall matter distribution, as opposed to significant evolution in the galaxy-halo relation. Durkalec et al. (2015b) measure the galaxy bias from galaxy clustering at $z \sim 3$ in the COSMOS field in VUDS to be $\sim$2.6 , which appears consistent with what one could expect extrapolating the bias measurements found in this chapter to higher redshifts.

Measurements of the satellite fraction (fig. 3.13) are consistent with McCracken et al. (2015), decreasing with redshift (as the higher mass halos within which most satellites reside would be have yet to form). As expected, there are more satellites at low redshifts; at $z \sim 1.5$ only around 5% of galaxies are satellites, which rises to around 20% by $z \sim 0.65$. The satellite fraction also begins to drop off slightly at stellar masses above $\sim 10^{10.5} M_{\odot}$, as these galaxies are only formed as centrals in higher mass halos, and the even more massive halos within which they would be satellites are very rare.

The measurements of r_0 (figure 3.14) are also qualitatively similar to McCracken et al. (2015), although as with the halo masses I do not yet see evidence for the upturn at higher masses. The lowest redshift bin is offset relative to the others, but this is to be expected, as it was also offset to higher halo mass values (figure 3.6), and even for just fixed halo masses I would expect some evolution in r_0 as the correlation function of dark matter increases towards lower redshift at fixed radii, independently of the halo occupation.

3.5.6 Stellar Mass to Halo Mass Ratio

The stellar mass to halo mass ratio (SMHR) is the total stellar mass in a halo (e.g. the stellar masses of all the galaxies in a halo summed) divided by the halo mass, and can be thought of as a measure of the star-formation and galaxy accretion history of a halo, or the global star accumulation efficiency for the whole halo. Evidence from the literature

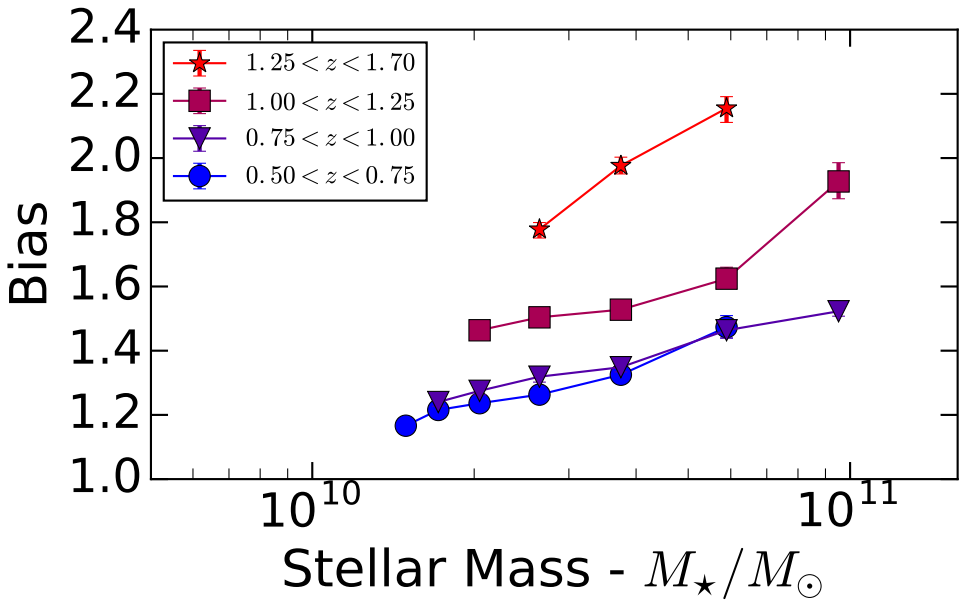


Figure 3.12: Galaxy bias as a function of stellar mass at the four redshift bins denoted in the legend. The bias increases with stellar mass, and also increases with redshift.

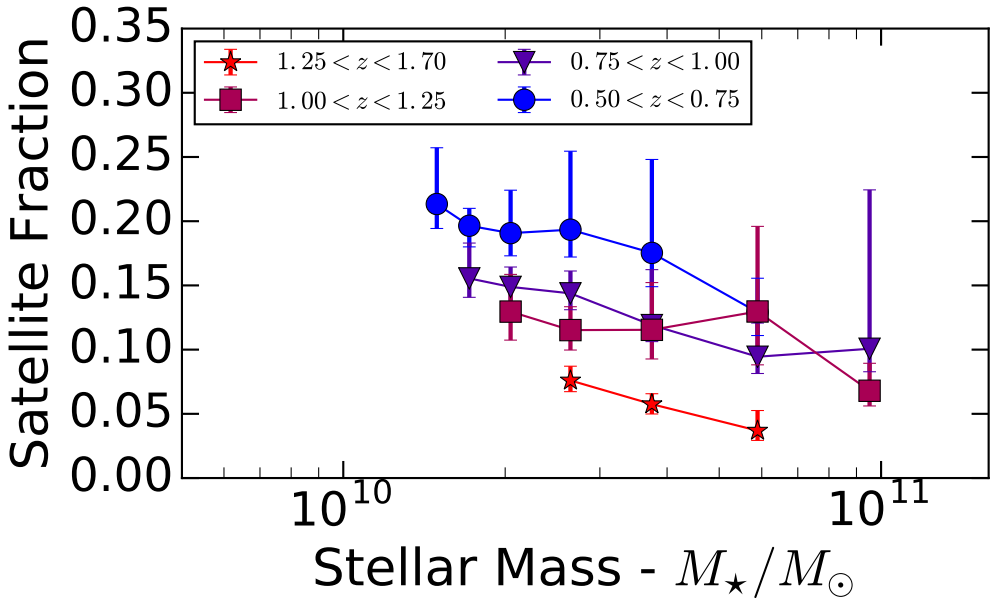


Figure 3.13: The satellite fraction as a function of stellar mass at the four redshift bins denoted in the legend. The satellite fraction is flat, tailing off at high masses, with more satellites at low z as expected.

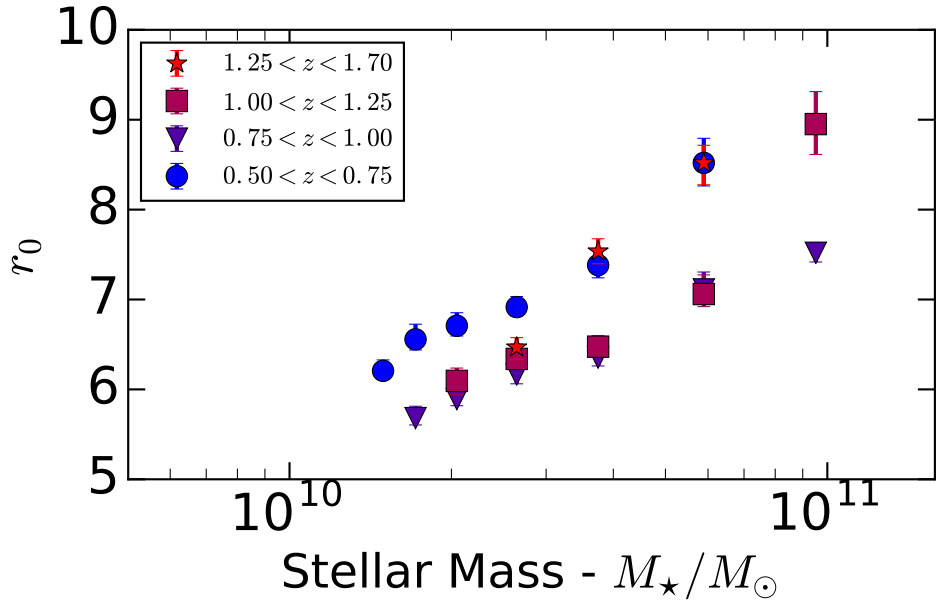


Figure 3.14: Evolution of r_0 with stellar mass. Colour represents redshift bin as denoted in the legend.

suggests it has a peak at halo masses of $10^{11.8} - 10^{12.4} M_\odot$ (see fig. 11 in McCracken et al., 2015, highlighting debate in the literature about possible redshift dependence), and is often modelled with a double power law as in Yang et al. (2003). I estimate the SMHR by simply integrating the best-fit HOD models (broken into central and satellite contributions as in Coupon et al. 2015) and bootstrapping the errors (figure 3.15). It can also be estimated by analytic inversion if the HOD model is fitted globally (e.g. HOD parameters are expressed as functions of stellar mass, as in Coupon et al. 2015) or abundance matching techniques with an N-body simulation (Kravtsov et al. 2004, Vale & Ostriker 2006 and Conroy et al. 2006), or from lensing measurements (as in Hudson et al. 2014). Simply integrating the HODs suffers systematics in that it can only underestimate the total stellar mass (as lower-mass galaxies are not included as the survey can not detect them and it has an artificial upper limit on galaxy mass because the massive galaxies are too rare to make clustering measurements). The error bars are also underestimated as I bootstrap each of the HOD models independently; in practice there is moderate covariance between the models as galaxies appear in multiple correlation functions. However it has the advantage of simplicity

and does not make extrapolations to galaxies the survey cannot study - either because of flux or volume limitations. The measurement of the SMHR rise from very low values at low halo masses (where halos only host a galaxy with a low probability), to reach a peak of $M_\star/M_h \sim 10^{-1.9}$ at a halo mass of $\sim 2 \times 10^{12} M_\odot$, in a regime where the central galaxy is much more massive than any satellites. It then declines to a local minimum at $\sim 3{\times}10^{13} M_\star/M_\odot$, where the transition from most stellar mass being in the central galaxy, to most being in satellites, occurs. Subsequently the number of satellites grows as a power law, and the SMHR grows again in the regime of clusters of hundreds of galaxies. This picture is in qualitative agreement (peak and local minimum before a power-law for ultra-massive halos) with Coupon et al. (2015) who conclude that including satellites in the SMHR can boost its value by an order of magnitude, which agrees with my findings here. Durkalec et al. (2015a) fit HOD models to the projected correlation function in VUDS and measure the SMHR at $z \sim 3$, finding the SMHR reaches the slightly higher value of $M_\star/M_h \sim 10^{-1.6}$. This could represent slight evolution, although it is very hard to make direct comparisons, given difficulties in making consistent comparisons of stellar mass estimates at different redshifts etc. The values of halo mass for peak SMHR found in this work are consistent with most of the literature values, with weak/no strong redshift dependence (figure 3.16). Note that not all techniques for calculating the SMHR are equivalent. Different definitions of halo mass aside (M_{vir} vs. M_{200} etc.), some authors effectively quote the central to halo mass, some (as I do here) quote the sum of the stellar mass of all the galaxies in the halo. McCracken et al. (2015) use median stellar mass to halo mass $M_{\rm min}$. I, following Coupon et al. (2012), use threshold stellar mass to $M_{\rm min}$, which explains why the SMHR measurements are consistently lower than McCracken et al. I can be moderately confident in the measurements around the peak, as it is in the range of halo masses probed by the HOD analysis (see fig. 3.6), but the estimates far outside this range should be approached with some caution. In particular, the apparent redshift dependence at low stellar masses is an unphysical consequence of not probing to lower stellar masses at high redshifts, and the central galaxy mass to halo mass ratio at high halo masses is unrealistically shallow as I cannot measure clustering for the most massive galaxies.

I also show in figure 3.15 the line $M_\star/M_{\rm halo} = \Omega_b/\Omega_{\rm DM}$ (using $\Omega_b = 0.049$, Planck

Collaboration, 2016), showing the ratio of baryonic to dark matter for the whole Universe. The measurements from VIDEO are safely under the line (e.g. I do not have more stellar mass than total baryons!) and indeed show that only a small fraction of baryons are in galaxies, as expected, Fukugita et al. (1998). I also plot the $z = 0$ line for the mass estimated to be in stars (6% of baryons, from Fukugita & Peebles, 2004). Although just an average, I see that around the peak of the SMHR, where the main contribution to stellar mass is, the measurements are very close to $\Omega_\star/\Omega_{\rm DM}$, consistent with most, but not all stellar mass seen today having assembled at $z \sim 1$.

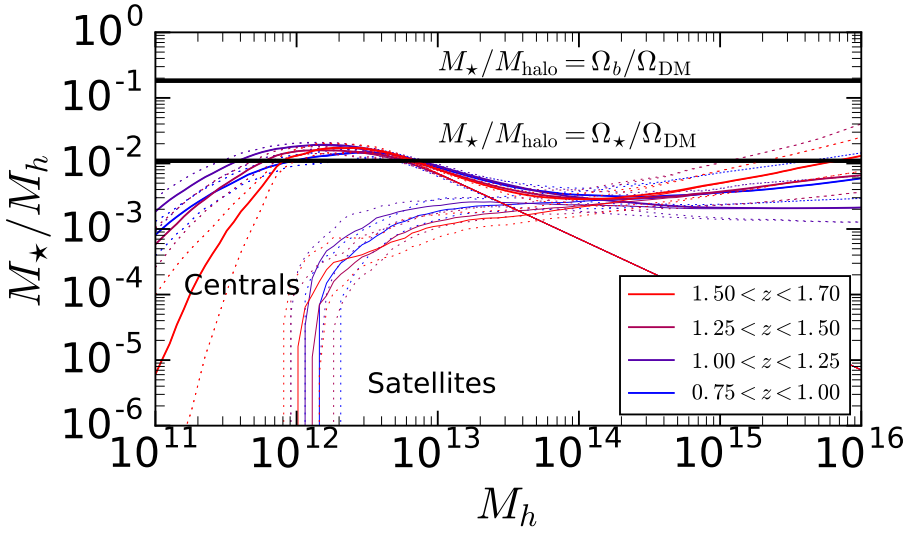


Figure 3.15: The stellar-mass halo ratio for different redshifts, divided into contribution from central galaxy and satellite galaxy. I plot for comparison $M_\star/M_{\rm halo} = (\Omega_b/\Omega_{\rm DM})$, and $M_\star/M_{\rm halo} = (\Omega_\star/\Omega_{\rm DM})$ from Fukugita & Peebles (2004)

It is perhaps surprising that little evolution is seen in the connection between stellar mass in galaxies and their host halos over cosmic time, particularly when other properties of galaxies are known to vary dramatically over the same epoch. For example, global comoving star formation density is well known to drop by around half an order of magnitude from $z \sim 2$ to $z \sim 0.5$ as galaxies increasingly have less gas to form new stars from (see Madau & Dickinson, 2014 for a review). Similarly, the (potentially associated) super massive black hole accretion density also drops by up to an order of magnitude (e.g. Hirschmann et al., 2014). Morphological properties (e.g. Sersic indices) of galaxies are believed to be relatively

stable from the local Universe to $z \sim 1$ (Cassata et al., 2007), but are typically observed to be dramatically different by $z \sim 2$ onwards (Lee et al., 2013). All these additional properties are extremely important in understanding galaxy evolution and the galaxy-halo connection.

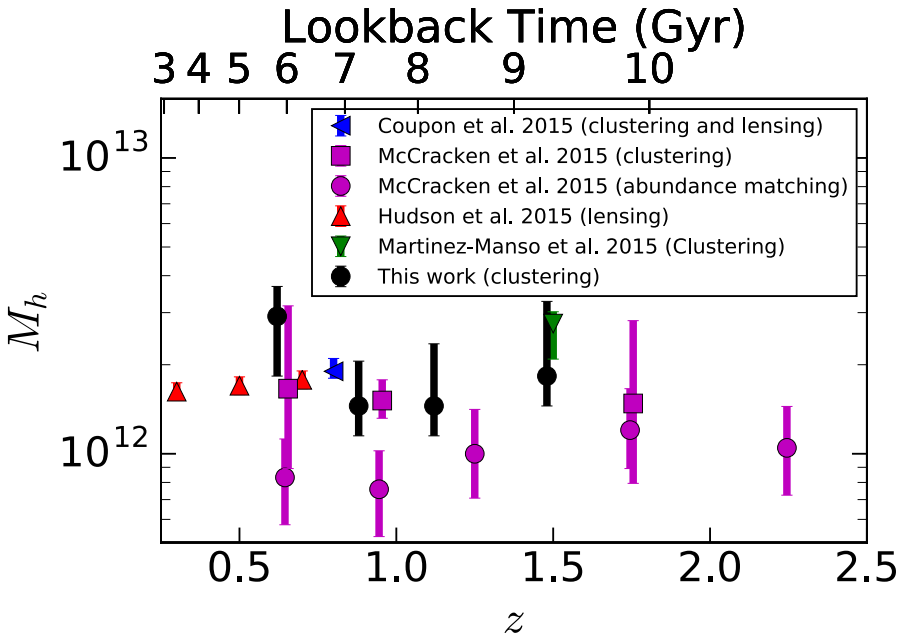


Figure 3.16: The redshift dependence of the peak of the SMHR from the data compared with several values from the literature.

3.6 Conclusions

In this Chapter I have used data from the VIDEO survey to investigate the galaxy-halo relation using 10-band photometric redshifts and stellar mass estimates. In particular I have studied the clustering of galaxies with the two point correlation function up to $z \sim 1.7$, using the Parzen-Rosenblatt estimator to calculate the correlation function in a novel way, without angular space binning and show it to be consistent with binned methods. Then a HOD analysis of the galaxy clustering was performed to give information about how galaxies occupy halos over cosmic time, as well to derive standard properties such as the bias of galaxies, and their satellite fraction. On the whole the data was found to be in good agreement with other surveys and clustering analyses, in particular the closely related UltraVISTA survey, another public VISTA survey currently at similar depths and breadths

to VIDEO, which in subsequent data releases will get deeper as VIDEO gets wider.

I see no substantial change in the occupation relations over time; all changes are driven by the change in the halo population. Typical halo mass increases with galaxy mass, and the ratio between the halo being sufficiently massive for one galaxy, to two, is around 15, suggesting that is the typical mass ratio between a halo of a given mass, and a halo massive enough to have enough substructure to have a sub-halo of that given mass. The power law relation for the number of satellites in a halo is ~ 1, and the scatter in halo mass to galaxy mass is broadly consistent with Tully-Fisher measurements at similar redshifts. I found bias increases with stellar mass, as galaxies are found in more massive, more biased halos, and decreases with time, as the halos trace the large scale dark matter distribution more accurately. The satellite fraction drops at high redshifts as the more massive halo within which satellites are found have not yet collapsed, and at high stellar masses as the super-massive halos within which high mass galaxies could be satellites are extraordinarily rare. Finally, my estimate of the stellar mass to halo mass ratio, although limited by the range of masses VIDEO currently probes, is in reasonable agreement with other studies, with a peak at a halo mass of around $2 \times 10^{12} M_{\odot}$ that is approximately constant in redshift.

UltraVISTA and this Chapter using VIDEO currently probe similar parts of parameter space but will start to diverge in future data releases. VIDEO-UltraVISTA complementarity is key - UltraVISTA gives a single instance of structure, I present another here, and future VIDEO results will subsequently give many more. UltraVISTA DR2 will probe several orders of magnitude deeper in the same field in all their near infra-red bands, allowing extension of their analysis to $z > 4$, and to lower stellar masses. Future work in subsequent VIDEO releases (when deeper optical data is available) will extend to larger areas over three separate fields (eventually 12 deg^2 in total), reducing uncertainty on measurements on the parameters reported in this paper, extending the angular scales probed by a factor $\sim$3 (allowing a better constraints on both the 1-halo and 2-halo terms), extending to more massive galaxies (allowing better analysis of the 'kink' in halo mass at high stellar masses), and giving an initial measure of cosmic variance by comparing results between the three fields.

Chapter 4

Environmental Quenching and Galactic Conformity in the Galaxy Cross-Correlation Signal

This Chapter closely follows Hatfield & Jarvis (2016, submitted).

4.1 Introduction

Galaxy colours have long been observed to be bimodal, with a 'red sequence' and 'blue cloud', with the 'green valley' separating them (e.g. Bell et al., 2004, see section 1.3.4). Extensive work interpreting these galaxy colours has allowed a connection to be made to physical properties of the galaxies e.g. the stellar populations that make up the galaxies, their stellar mass and star formation rates (Bruzual & Charlot, 2003 and many others). It is established that the red sequence corresponds to a passive 'red and dead' population of massive galaxies, where star formation has ceased, and that the blue cloud corresponds to typically less massive galaxies that are still actively forming stars (Strateva et al., 2001) - this observational result leads naturally to the question of what causes this transition (referred to as 'quenching'), and where and when does it occur for which galaxies? Temporally, as discussed in section 1.3.5, global star formation is known to peak around 6-7 billion years ago, and to have dropped off since (Madau et al., 1996, 1998; see Madau & Dickinson, 2014

for a review). However, spatially, it is also known that, beyond redshift and stellar mass, the environment a galaxy is found in can have a large effect on whether it is quenched or not, the so-called 'colour-density' relation, that passive, redder galaxies are typically found in denser regions of the Universe. This holds even after controlling for the fact that more massive galaxies are biased towards denser regions, and is closely related to the 'morphology-density' relation, that these passive galaxies in dense regions are typically elliptical, and the star forming galaxies in less dense regions are more likely to be spirals. This phenomenon was first investigated by Oemler, Augustus (1974); Davis & Geller (1976); Dressler (1980) and others, and now enjoys extremely robust and well studied measurements in the local Universe e.g. in the Sloan Digital Sky Survey (SDSS; Ball et al., 2007) and in the Two Degree Field Galaxy Redshift Survey (2dFGRS; Balogh et al., 2004).

What processes might give rise to these relations in the galaxy population? Key processes in galaxy evolution are often classified into 'nature' and 'nurture' effects, e.g. internal processes such as cooling and feedback versus interactions with other galaxies and the local environment - often a variety of processes are invoked to explain environmental-based observations. For example, many mechanisms involving removing the cool gas needed for star formation have been proposed to account for the observed quiescence in satellite galaxies around central massive galaxies. They range from strangulation (tidal effects from the gravitational potential allows the gas in the satellite to leave) and ram pressure stripping (removal of gas by 'winds' in the hot intra-cluster medium, Gunn & Gott, 1972) to harassment (flybys from other galaxies, Farouki & Shapiro, 1981 and Moore et al., 1998) - see Hirschmann et al. (2014) for more details. The mass of the dark matter halo is also believed to have a key role in these processes; Hartley et al. (2012) suggest that because only a modest energy input is needed to keep the gas in the halo hot, there should be a critical dark matter halo mass that divides those that host passive and those that host star forming galaxies. A study by Peng et al. (2010) using SDSS and zCOSMOS (Scoville et al., 2007) takes an empirical approach to modelling the build up of stellar mass in galaxies and concludes that quenching effects can be strongly distinguished into environmental quenching (e.g. satellite galaxy infall processes), merger quenching and mass quenching, an unknown process with strength proportional to the star formation rate which the authors speculate could involve

active galactic nuclei (AGN) feedback. Understanding what gives rise to these independent 'environmental' quenching and 'mass' quenching effects is an ongoing challenge.

Although host halo mass is known to be important, it is also becoming apparent that it is not the only factor in environmental quenching. Multiple studies have reported correlations in the properties of nearby galaxies, even after accounting for halo mass, an effect now known as galactic conformity (e.g. Hearin et al., 2016a). Galactic conformity typically manifests as correlations in the sSFR of galaxies spatially close e.g. a tendency for galaxies close together to either all be passive or all star-forming. More specifically, conformity can be broken down into 1-halo and 2-halo conformity (in analogy with the 1-halo and 2-halo terms in halo occupation distribution modelling). 1-halo conformity describes correlations within a halo, typically scales less than a megaparsec (see Weinmann et al., 2006) and 2-halo conformity describes correlations between nearby halos, typically on scales of more than a megaparsec (Kauffmann et al., 2013). 1-halo conformity, although not perfectly understood, does not contradict the basic picture of dark matter halos as isolated environments; it is possible to imagine that it could be explained through galaxy interactions, AGN, and other mechanisms within the halo. 2-halo conformity is less straightforward and does not permit such a simple explanation for why correlations exist on such large scales, violating the usual assumption of the halo model that everything inside a halo only depends on the halo mass. A small 2-halo conformity would be expected from the fact that massive halos are typically clustered, and more massive halos typically contain more quenched objects, leading to objects being more passive on scales larger than their individual halos - but conformity appears to be larger than expected from this effect. Hearin et al. (2016a) suggest 2-halo conformity could be explained through large-scale tidal environments affecting individual halos and/or as a consequence of assembly bias, where the assembly history of a halo has an effect on its bias and also its contents. The most simple manifestation of halo bias is the result that at fixed halo mass, halos that were assembled earlier are more highly biased. Hearin et al. (2016a) suggest that the halos that are assembled earlier will contain older galaxies that are more likely to have used up their gas reservoir; these galaxies will be more highly clustered and this is observed as 2-halo conformity.

Environmental quenching and the colour-density relation are typically studied either in

the context of group and cluster catalogues or nearest-neighbour type density measurements. Conformity has typically been observed thus far through satellite density profiles and similar techniques e.g. Hartley et al. (2015). Both the colour-density relation and conformity make up part of the statistical distribution of stellar mass and star formation and thus should be detectable in the correlation function. In Chapter 3 I used the VISTA (Visible and Infrared Survey Telescope for Astronomy) Deep Extragalactic Observations (VIDEO) survey (Jarvis et al., 2013) to model the two-point correlation function and study the galaxy-halo relation. In this Chapter I use the same observational selection criteria and cuts, and a similar approach to the calculation of the correlation function, to investigate environmental effects. I seek to investigate this by looking at the cross-correlation signal between higher-mass and lower-mass galaxies, and then varying the star formation rates of the samples.

This chapter is organised as follows: first I summarise the sample selection made in Chapter 3, and discuss how I measure the cross-correlation function. I then measure the cross-correlation function for a series of sub-samples split by mass, redshift and star formation rate in the survey. Finally I interpret my results in terms of environmental quenching and conformity by incorporating a quenching model into the Halo Occupation Distribution (HOD) framework.

4.2 Observations

In this chapter I use the same catalogue described in Chapter 3. I briefly summarise the observations and sample selection here.

The catalogue was constructed by combining the NIR VIDEO observations (Jarvis et al., 2013) with optical data from CFHTLS D1 tile (Ilbert et al., 2006; Gwyn, 2012). This gives deep photometric imaging across 10 bands over 1 deg^2. The sources in the images were identified using SExtractor, (Bertin & Arnouts, 1996), and photometric redshifts, stellar masses and star formation rates used in this study were calculated using LePHARE (Arnouts et al., 1999; Ilbert et al., 2006). I then removed sources in regions effected by excess noise and bright stars with a mask, a K_s >23.5 magnitude cut, and colour cut around a stellar locus, following the approach of Baldry et al. (2010), to remove stars (see

figure 3.2). VIDEO has a 90 percent completeness at this depth (Jarvis et al., 2013) and McAlpine et al. (2012) estimate this colour cut leaves stars contributing less than 5 per cent of the sample.

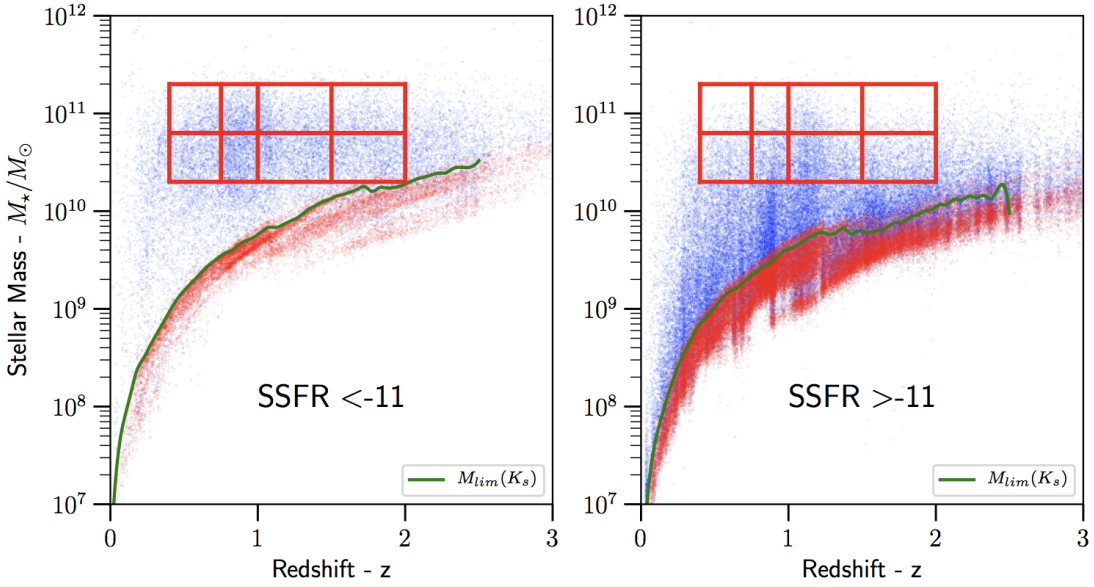


Figure 4.1: The mass and redshift of galaxies, considered after application of the magnitude cut, star exclusion and mask, are shown here in blue. The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of $K_s < 23.5$, and the green curve the implied 90 per cent stellar mass completeness limit, following the approach of Johnston et al. (2015). The red boxes illustrate the redshift and stellar mass selected sub-samples that I consider in section 4.4, with the left plot corresponding to the passive sample, and the right hand plot the star-forming sample.

Figure 4.1 shows the final sample in stellar mass-redshift space broken down into passive ($\log[\mathrm{sSFR/yr}^{-1}] < -11$) and starforming galaxies ($\log[\mathrm{sSFR/yr}^{-1}] > -11$), with the bold red lines delineating my sub-samples. Above the green curve is the part of the parameter space within which there is high confidence of completeness (see Johnston et al., 2015 for a discussion of the calculation of completeness limits). Conversely the red dots represent objects where incompleteness may be a factor - I only probe areas with high completeness.

Figure 4.2 shows a scatter plot of stellar mass versus specific star formation rate for the four redshift bins that we use throughout this chapter, and the subsamples used. Table 4.1 summarises the number counts and associated passive fractions. Star formation rates are derived from the templates used to construct the photometric redshifts (that is to say each

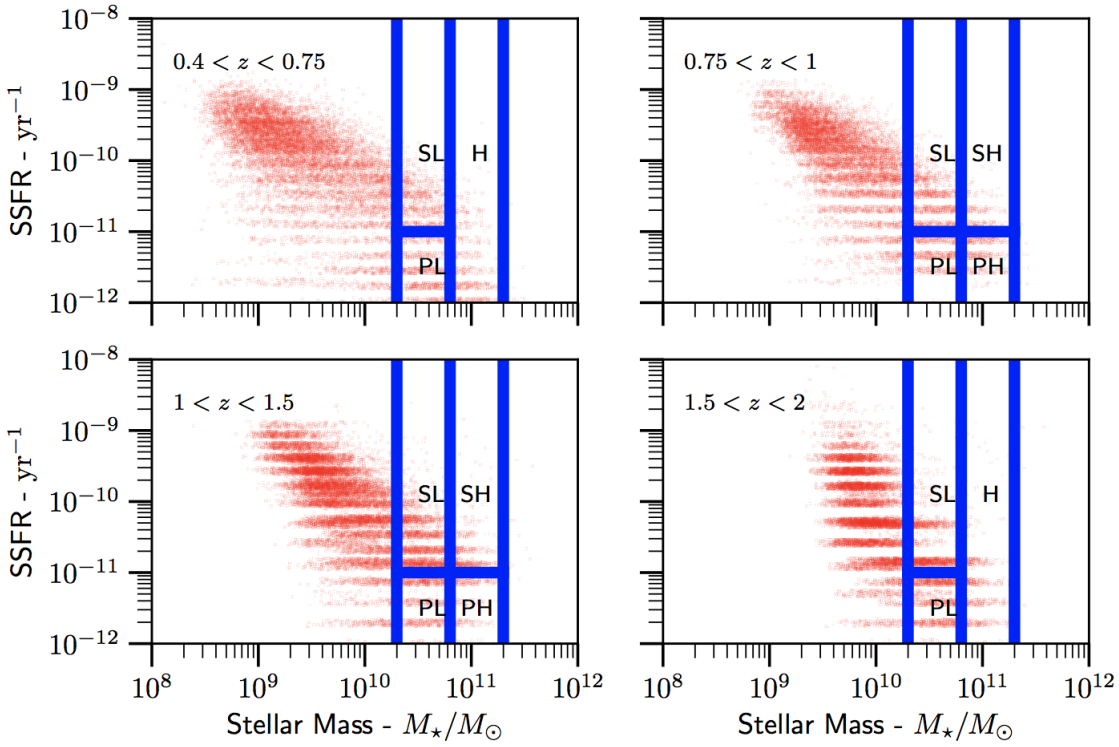


Figure 4.2: The mass and sSFR of galaxies in the final sample, in the four redshifts considered in section 4.4. The blue lines delineate the various samples. The initials indicate which section corresponds to each sample: 'P'=passive, 'S'=star forming, 'H'=high mass, 'L'=low mass.

template was constructed with a star formation history, and the star formation rate is that of the best fitting template). The red sequence and blue cloud/main sequence bimodality is just about visible in figure 4.2, with lower mass galaxies having higher specific star formation rates in the main sequence. The sSFR values not taking certain values is an unphysical result of the fact that star formation rates are calculated from a discrete set of templates. See Jarvis et al. (2013) for a more in depth description of the templates used in the photometric redshift fitting. Our results for the $M_\star -$ sSFR plane are qualitatively similar to results in the literature, both in median relationship, and in scatter - see Rodighiero et al. (2010) for analogous results at similar redshifts from Herschel Space Observatory data, and Ciambur et al. (2013) for comparable results in UDS, SDSS and the Great Observatories Origins Deep Survey (GOODS) data. Johnston et al. (2015) gives a more in depth analysis and discussion of the $M_\star -$ sSFR in the VIDEO survey.

Table 4.1: Number of galaxies for each redshift, stellar mass, and star formation rate bin, with the associated passive fraction for the low mass galaxies. H denotes the high mass ($10.9 < \log_{10}(M_\star/\mathrm{M}_\odot) < 11.4$) galaxies, PL the passive low mass galaxies, and SL the star forming low mass galaxies ($10.4 < \log_{10}(M_\star/\mathrm{M}_\odot) < 10.9$, $\log_{10} sSFR \lesssim -11$). Note that these counts are actually the sums of the weights of the galaxies in each redshift bin.

Redshift Range	N_g (H)	N_g (PL)	N_g (SL)	f_p
$0.40 < z < 0.75$	350	1193	628	0.655
$0.75 < z < 1.00$	524	1622	860	0.653
$1.00 < z < 1.50$	748	2171	1880	0.535
$1.50 < z < 2.00$	400	1670	975	0.631

As can be seen from Table 4.1, the passive fraction is ~ 0.65 at the lower redshifts, drops slightly in the $1 < z < 1.5$ redshift bin and then rises again in the $1.5 < z < 2$ bin. This is slightly unexpected - typically one would expect the passive fraction at fixed stellar mass to drop with redshift (e.g. Tinker et al., 2013). I suggest that this high passive fraction at high redshift could be due to a) the difficulties of estimating star formation rates in dusty galaxies e.g. some of the 'passive' galaxies could actually be reddened dusty star forming galaxies and b) potentially cosmic variance on the moderately small samples sizes (see Fig.7 in Johnston et al., 2015). Note however that number counts do not impact on the correlation function as long as the probability of being observed versus not being observed has no spatial dependence.

The star formation properties of galaxies in VIDEO have been explored in other works (Zwart et al., 2014; Johnston et al., 2015). However these works have not studied *where* the star formation is occurring.

4.3 The Angular Cross Correlation Function

4.3.1 Background

As discussed in Chapters 2 and 3, angular correlation functions are a useful probe of the large-scale spatial distribution of the galaxy population. In Chapter 3 I considered the *auto*-correlation function, the excess probability above Poisson of finding two galaxies of a single population at a given angular separation. In this Chapter I consider the *cross*-correlation function, the excess probability above Poisson of finding two galaxies each from a different

population at a given angular separation.

The conventional way to estimate the auto correlation function $\omega(\theta)$ is with the Landy & Szalay (1993) estimator (as discussed in Chapter 2, and used in Chapter 3), which is based on calculating $DD(\theta)$, the normalised number of galaxies at a given separation in the real data, $RR(\theta)$, the corresponding figure for a synthetic catalogue of random galaxies identical to the data catalogue in every way (i.e. occupying the same field) except position, and $DR(\theta)$, the number of galaxy to synthetic point pairs. Szapudi & Szalay (1998) generalise this to the cross correlation function with:

$$\omega(\theta) = \frac{D_1D_2 - D_1R - D_2R + RR}{RR}, \tag{4.1}$$

where $D_1D_2(\theta)$ is the number of 'population 1' to 'population 2' pairs of galaxies at a separation θ, $D_1R(\theta)$ is the number of 'population 1' to 'random' pairs etc.

The cross-correlation function can be used to demonstrate a physical association between two phenomena. Cross-correlation of the CMB lensing potential with high redshift active galaxies has proved fruitful in showing these galaxies live in massive halos (e.g. Sherwin et al., 2012; Geach et al., 2013; Allison et al., 2015). Conversely the cross-correlation function can also be a useful check that two distributions are not correlated when they are not expected to be, in which case the cross-correlation should be consistent with zero. Coupon et al. (2015) for example cross-correlate galaxies at different redshifts to validate their photometric redshifts - different redshift bins should be causally separated parts of space and hence should have negligible cross-correlation. A non-zero cross correlation would correspond to cross contamination between the two bins.

The cross-correlation function has also been shown to have great potential in the context of the multi-tracer technique (Seljak, 2009), in which galaxy samples which trace the underlying matter distribution in different ways are used simultaneously to constrain cosmological parameters with a greatly reduced contribution to the error budget from cosmic variance. Cross-correlations can also be used to make inferences about populations with no redshift information (e.g. un-matched radio sources) when correlated with a population that does (e.g. Fine et al., 2015).

The literature is rich with approaches to modelling the auto-correlation function. To a fairly high degree of precision, a power-law can be fitted with a gradient of ~ -0.8. More recently, Halo Occupation Distribution (HOD; Cooray & Sheth, 2002) models of the 2-point correlation function have seen great success in galaxy evolution studies, and give physical results in agreement with other methods (e.g. Coupon et al., 2015; Chiu et al., 2016) - see Chapter 3.

On large scales, viewing galaxies as biased tracers of the dark matter distribution in the linear regime, the cross correlation function can be easily modelled. The spatial auto-correlation functions of two galaxy samples in the linear regime are $\xi_{\rm galAA} = b_{\rm A}^2 \times \xi_{\rm DM}$ and $\xi_{\rm galBB} = b_{\rm B}^2 \times \xi_{\rm DM}$, where $b_{\rm A}$ and $b_{\rm B}$ are the galaxy biases of the samples, and $\xi_{\rm DM}$ is the dark matter correlation function. The cross-correlation between the two samples on linear scales is then $\xi_{\rm galAB} = b_{\rm A} \times b_{\rm B} \times \xi_{\rm DM}$ e.g. the geometric mean. On these scales the cross-correlation function does not provide any additional information than that contained within the two auto-correlation functions e.g. $\xi_{\rm galAB}^2 = \xi_{\rm galAA} \times \xi_{\rm galBB}$. This can be exploited to estimate the bias of samples that are too small to measure the auto-correlation function (e.g. Lindsay et al., 2014 cross-correlate radio galaxies and IR galaxies, Knobel et al., 2012 cross-correlate galaxies and galaxy-groups).

Simon et al. (2009) show how to extend the HOD approach to the cross-correlation function with the joint halo occupation distribution (JHOD) model. Within conventional HOD, the 1-halo term is constructed by finding the contribution to the clustering from each halo mass, and weighting by the halo mass function (HMF). The contribution from each halo mass is the convolution of the halo profile with itself, weighted by $\langle N^2 \rangle$, where N is the average number of galaxies in a halo of that mass. For a cross-correlation function, this weighting is instead replaced with $\langle N_A N_B \rangle$, where N_A and N_B are the average numbers of galaxies of type A and B in a halo of that mass. Effectively the two auto-correlation functions provide information about $\langle N_A^2 \rangle$ and $\langle N_B^2 \rangle$, and the cross correlation gives you information about $\langle N_A N_B \rangle$ i.e. on small scales the cross-correlation *does* provide information not contained within the two auto-correlation functions ($\xi_{\rm galAB}^2 \neq \xi_{\rm galAA} \times \xi_{\rm galBB}$). In particular it gives the *covariance* of the two occupations. This can be scaled to lie in the range [-1,1] e.g. for a given halo mass, members of the two samples are never found together

(-1), occupy the halos independently (0), or are always found together (+1). Its impact is on the 1-halo term of the cross-correlation function, to enhance it when galaxies are always found in the same halo, and dampen it when the galaxies are rarely found in the same halo.

Simon et al. (2009) fitted the auto and cross correlation functions for red and blue galaxies simultaneously. They also investigated the effect of allowing red and blue galaxies to lie on different profiles, in which case the 1-halo cross-correlation function is proportional to the convolution of these different halo profiles. These two approaches ('JHOD' introducing covariance into how the galaxies occupy the halos, and altering the profiles the different samples followed) have the same qualitative effect of modelling different galaxy types having an affinity (or not) to be nearby, but are mathematically distinct; Simon et al. (2009) conclude that the two models were broadly degenerate for the levels of precision explored in their work.

4.3.2 Application to Environmental Quenching and Conformity

As discussed in section 4.1, the star formation rate of galaxies is believed to be tied to the environment within which they live. Woo et al. (2013) and Gabor & Dave (2014) track how the quenched fraction changes as a function of radial location in the halo and the halo mass. This, as part of the description of the spatial distribution of galaxies, must play a role in determining the correlation function, but is not typically included in modelling of correlation functions. Most HOD studies typically just make measurements as a function of luminosity/stellar mass (Coupon et al., 2015; Hatfield et al., 2016). Some (e.g. Coupon et al., 2012; McCracken et al., 2015) model the auto-correlation of the whole galaxy population, and then the passive galaxy population, finding that the passive galaxies are typically found in higher mass galaxies, but is not typically linked to measurements of the passive fraction etc. Related to the idea of environmental quenching, conformity refers to the observation that nearby galaxies have correlated star formation rates. Again, this is an observation about spatial distributions of galaxies, so correlation functions seem well suited to describe them.

In this Chapter I suggest a potentially valuable way of observing environmental effects (including conformity) in what I refer to here as the 'environment tracer method'. I take a

'tracer' population, which in some sense is believed to trace, or even cause, some particular environment, and a 'follower' population, which has its properties predominantly decided by the environment it finds itself in. In this Chapter I choose the tracer population to be massive galaxies, as only these have the ability to have a large impact on their environment, and lower mass galaxies as the follower population. I then measure the cross-correlation between the tracer and follower populations as a function of their properties e.g. the specific star formation rate of each population. This allows us to see if the properties of the follower population are affected by the properties of the tracer population - comparatively lower power in the cross-correlation means the follower population is inhibited by the tracer population, comparatively high power in the cross-correlation means the follower population is preferentially found near the tracer population. I note that the cross-correlation function is formally symmetric between the two populations; the distinction between tracer and follower population is purely one made for ease of discussion and comparison.

One can investigate for example how other measurements of conformity translate into the environment tracer method. In figure 10 of Weinmann et al. (2006), the fraction of satellite galaxies in a halo found to be late or early is measured as a function of halo mass, for a variety of samples. They found that, at fixed halo mass, the late (early) fraction increased when the central itself was late (early). The natural equivalent to look at this through correlation functions is to see the impact on the cross correlation of massive galaxies and less massive galaxies, when the massive sample is split by type (early/late). One would expect a reduction in the amplitude of the cross correlation from massive lates to low mass early types, and vice versa. In figure 2 of Kauffmann et al. (2013), the median sSFR of galaxies as a function of a distance from a massive central galaxy is measured. They find that the median sSFR increases up to radii of $\sim$ 4 Mpc when the sSFR of the central is increased. The environmental tracer method equivalent again is to make the central massive galaxies the tracer population and the other galaxies the follower population. One would then expect the cross-correlation function to be enhanced when the populations have correlated star formation rates.

4.4 Results

My approach is to measure galaxy interactions/the effect galaxies can have on each other through the environment tracer method. When I calculate RR for the estimation of the correlation function I use 500,000 random data points in the 1 deg^2. For the uncertainties on the correlation functions, as per Chapter 3, I use bootstrap resamplings to estimate the uncertainty by taking the 16th and 84th percentiles of the resampling. I calculate the correlation function as per Chapter 3, estimating the correlation function a) using the kernel smoothing method described in Chapter 2, b) using the weights system of Arnouts et al. (2002) to account for uncertainty in the redshift estimates, and c) treating the integral constraint as part of the model as per Beutler et al. (2011).

4.4.1 The High- to Low-mass Cross-Correlation Function

In this section I use the environment tracer method as described in section 4.3.2, with higher mass galaxies ($10.9 < \log_{10}(M_{min}/\mathrm{M}_\odot) < 11.4$) as the tracer population, and lower mass galaxies ($10.4 < \log_{10}(M_{min}/\mathrm{M}_\odot) < 10.9$) as the follower population. I then measure the cross correlation function between the tracer and follower populations, and sub-divide the follower population by star formation rate ($\log_{10} sSFR \lessgtr -11$). I do this for four redshift bins: $0.4 < z < 0.75$, $0.75 < z < 1.00$, $1.00 < z < 1.50$ and $1.50 < z < 2.00$. Redshift range choices were motivated by the desire to get roughly similar number of sources in each bin. The choice of $\log_{10} sSFR \lessgtr -11$ was motivated by a combination of a need to have a large number of galaxies in each sample, as well as literature results for the divider between passive and main sequence galaxies (e.g. Knobel et al., 2015 divide at $\log_{10} sSFR = -11.25$ at $M_\star = 10^{11} M_\odot$). I refer to samples as being 'passive' or 'starforming' based on this divider in this work, although I acknowledge that this divider could include the lower portion of the main sequence in the passive sample under some definitions (e.g. Choi et al., 2014 divide at $\log_{10} sSFR = -12.7$ at $M_\star = 10^{11} M_\odot$).

In figure 4.3 each subplot shows the cross-correlations for a different redshift bin. Each curve is the cross correlation between the 'tracer' massive galaxy sample, and a low mass 'follower' sample. The red curve shows the cross-correlation when the low-mass sample is

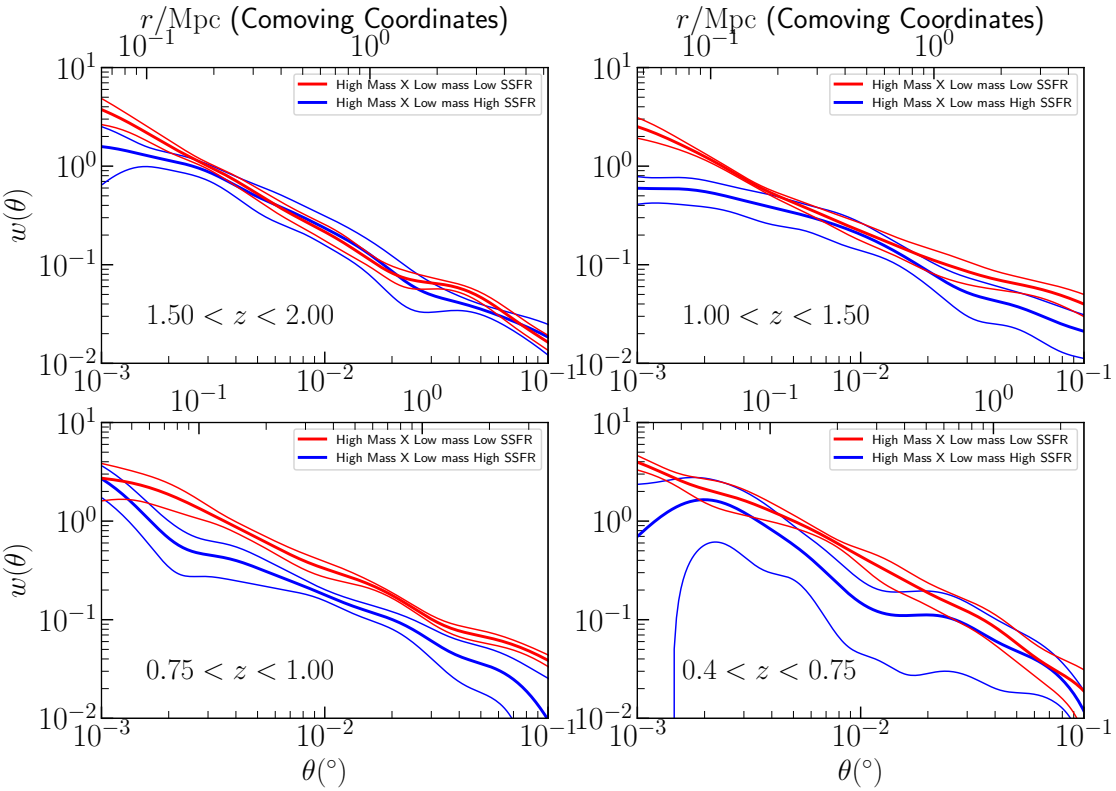


Figure 4.3: The cross-correlation function signal between low mass ($10.4 < \log_{10}(M_{min}/M_{\odot}) < 10.9$) and high mass ($10.9 < \log_{10}(M_{min}/M_{\odot}) < 11.4$) galaxies in VIDEO in four redshift bins. The low mass sample is selected to be passive ($\log_{10} sSFR < -11$) for the red curves, and star forming ($\log_{10} sSFR > -11$) for the blue curves. At the highest redshift, the two curves are very similar; the role of environment in determining star formation rate seems to be small. At $1 < z < 1.5$ the passive low-mass galaxies are more clustered around the massive galaxies, but only on small scales. In the $0.75 < z < 1$ and $0.4 < z < 0.75$ bins, the enhancement has reached the larger scales, and is substantial on all scales.

passive, and the blue curve star forming.

In the highest redshift bin ($1.5 < z < 2$), although the error bars are large, there is no significant difference between the two curves on all but the smallest scales. This indicates that, although there may be different number densities of the passive and star forming low mass samples, there is no significant difference in their geometric distribution around massive galaxies, both within the halo of the massive galaxy, and in the nearby halos i.e. there appears to be no environmental effect on star formation in the low mass galaxies, apart from the central ~ 0.1Mpc around the massive galaxy.

At the intermediate redshifts, $1 < z < 1.5$, there is substantial enhancement of the cross-correlation with the passive low-mass sample, compared with the star forming low-mass sample, out to a projected radius of 0.2 Mpc, roughly the length scale of the halos hosting the massive galaxy sample. I interpret this as a clear signal of quenching associated with the environment of the massive galaxy sample.

However for the lowest two redshift bins, I see the enhancement extends to all scales that can be probed with the 1 deg^2 field (albeit with large error bars in the lowest redshift range). By this stage environmental quenching has had such an effect that, although the average host halo mass was originally the same for passive and low-mass star-forming galaxies, they have been preferentially quenched in higher mass haloes, leaving the average halo mass higher for the passive tracer sample, imparting the higher bias evident in the higher two-halo term.

Note I have chosen to attempt to measure the cross-correlation at different redshifts for the same galaxy properties e.g. the galaxies have the same stellar mass and sSFRs at $1.5 < z < 2$ as $0.4 < z < 0.75$. This means I am not tracing the 'same system' over time, as a galaxy of $M_\star \sim 10^{11}\mathrm{M}_\odot$ at $z \sim 2$ will have a substantially larger mass at $z \sim 0.5$. The logical alternative to my approach would be to use evolution models motivated by simulations and theory to attempt to track the growth of the stellar mass of a galaxy over time. Similarly, I could have chosen to use a redshift dependent sSFR cut. I chose not to take this approach mainly on the grounds that the underlying physical process behind quenching being probed should be redshift independent e.g. statements like "a galaxy of star formation X has its star formation reduced by a factor of Y when it enters environment Z" should not depend on the time at which the event occurs. A given process may be more or less prevalent at different redshifts, but underlying physics of the processes should not change. Had I used a sSFR threshold that evolved with time it would be harder to make direct comparisons between different redshifts e.g. had I used $\log_{10} sSFR \leqslant -12$ at $z \sim 0.5$, and $\log_{10} sSFR \leqslant -11$ at $z \sim 1.5$ to divide between star-forming and passive galaxies, then I would be testing if a given environment can reduce the $\log_{10} sSFR$ below -11 at one time period, but below -12 at the more recent time period, so it would not be comparing like-to-like.

4.4.2 Selection by SFR in the Massive Sample

I now split the environment tracer sample by specific star formation rate. This is more challenging, as the star-forming fraction of the massive sample is much smaller, using the same divider of $\log_{10} sSFR \lessgtr -11$. I therefore only measure the effect in the two middle redshift bins ($0.75 < z < 1.00$ and $1.00 < z < 1.50$) to explore this effect, as the massive galaxies are almost entirely star forming at the highest redshift, and almost entirely passive at the lowest redshift.

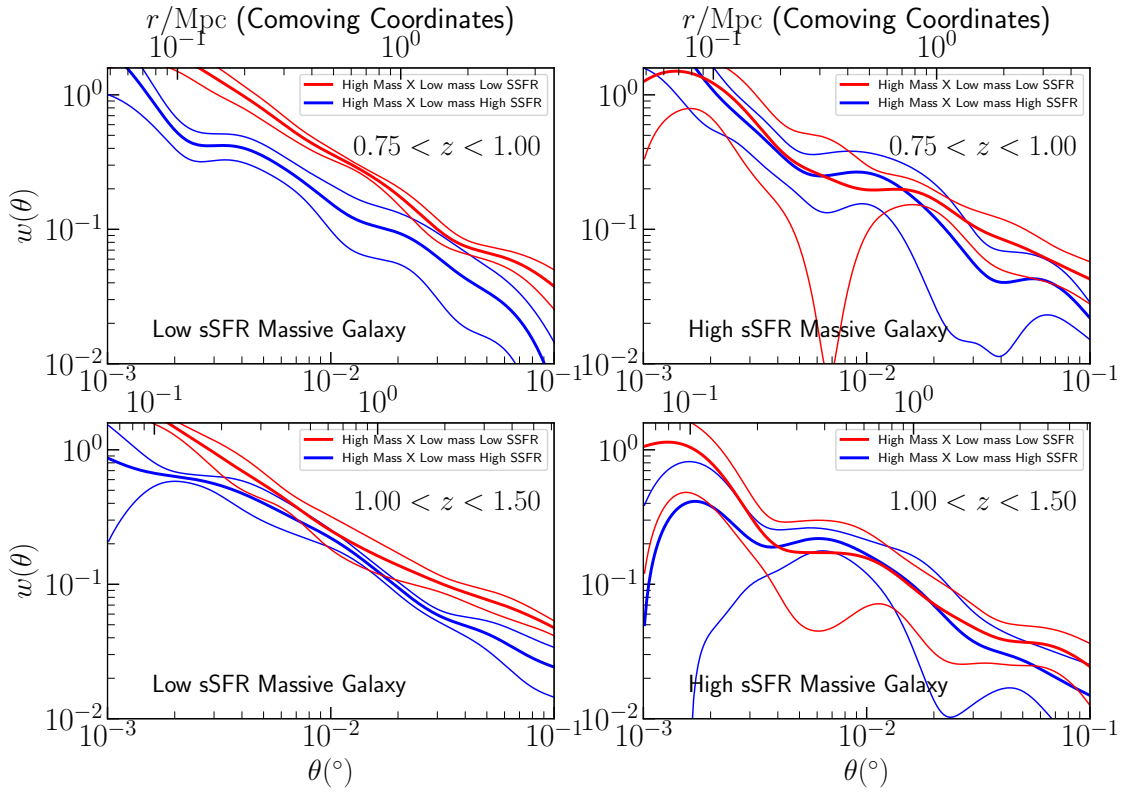


Figure 4.4: The cross-correlation function signal between low mass ($10.4 < \log_{10}(M_{min}/M_\odot) < 10.90$) and high mass ($10.9 < \log_{10}(M_{min}/M_\odot) < 11.4$) galaxies in VIDEO at two redshifts. Red denotes when the low mass galaxy is passive ($\log_{10} sSFR < -11$), blue denotes starforming ($\log_{10} sSFR > -11$). In the left two plots the massive galaxy is passive, and the right two the massive galaxy is starforming (with criterion $\log_{10} sSFR \lessgtr -11$).

In fig. 4.4, the top two panels are at $0.75 < z < 1.00$, and the bottom two sub-plots at $1.00 < z < 1.50$. Each subplot is the same as in Fig. 4.3 (e.g. the red corresponds to passive low mass sample and blue corresponds to the star forming, low mass sample), except there

is also sSFR dependency in the massive sample. In the left two panels the massive sample is selected to have a low sSFR, and in the right two panels, is selected to have high sSFR.

At both redshifts, when the massive central galaxy is passive, there is qualitatively similar behaviour to that presented in fig 4.3 - the passive massive sample have passive low-mass galaxies more clustered around them than star-forming low-mass galaxies. However there is essentially no enhancement when the massive galaxy is star-forming - lower-mass galaxies are equally clustered around them, suggesting that they have a similar passive fraction when near a massive star forming galaxy to the total population. In other words, at fixed central stellar mass, the lower mass satellites have lower star formation rates if the central is passive, to if the central is star forming. Furthermore, it appears that the size of the effect increases from $1.00 < z < 1.50$ to $0.75 < z < 1.00$.

This illustrates what conformity looks like in cross-correlation functions. Massive passive galaxies have lower mass passive galaxies much more clustered around them than lower mass star forming galaxies. Massive star-forming galaxies appear to have no star formation rate dependence for how satellites are clustered around them. In Section 4.4.3 I describe a way to advance our understanding of the role of the environment, while also controlling for halo mass, by incorporating these effects into the HOD formalism.

Note that this *is* a measurement of 2-halo conformity at $z \sim 1$, in the sense that there are correlations between star formation rates of galaxies on scales larger than individual halos. However it is *not* (yet) evidence that this is above and beyond what one would expect from the fact that massive halos are typically near to other massive halos and galaxies are typically more passive in massive halos. Further modelling is needed to investigate if there is a conformity effect suggestive of large scale quenched areas from assembly bias or tidal environmental effects on intermediate scales between individual haloes and the larger linear scales.

4.4.3 Modelling Quenching Effects Directly

Given these results, I seek a way of modelling them within the HOD framework. I note that the two innovations of Simon et al. (2009), allowing covariance of the red and blue galaxy occupation numbers, and allowing red and blue galaxies to trace different profiles,

are both ways of modelling red and blue galaxies 'not liking' being together; effectively an interaction term where galaxies are more likely to be passive when nearby to other passive galaxies - in other words conformity. My approach is to model the impact of this interaction on the correlation function from a quenching mechanism directly. I briefly discuss a toy model building on Simon et al. (2009) here.

4.4.3.1 Radial Dependence of Quenching Within a Halo

A reasonable fiducial model for the location dependence of the probability of a galaxy being quenched is that it either depends on radial distance from the centre of the halo, or it depends on distance to some other quenching source (a more massive galaxy, an AGN etc.). Three distinct scenarios (illustrated in figure 4.5) then arise in the context of my earlier example:

- Quenching depends on the separation of both galaxies, and both populations trace the halo profile
- Quenching depends on the radial distance to the centre of the halo, both populations trace the halo profile
- Massive galaxy/AGN sits at centre of the halo, impossible to distinguish between quenching coming from the massive object, or from radial distance to the centre of the halo

I denote the probability of being quenched in either of these scenarios as $Q(r)$, and use the parametrisation:

$$Q(r) = K + \frac{S-K}{2}\left(1 - \mathrm{erf}\left(\frac{\log_{10}(r) - log(r_Q)}{a}\right)\right), \tag{4.2}$$

with $0 < K < 1$, $0 < S < 1$, $0 < r_Q$, $0 < a$. This parametrisation is chosen to have galaxies quenched with probability S when they are at the source of the quenching, and quenched with probability K when they are far away. The length scale r_Q describes the scale of this transition, and a how sharp this transition is. Correspondingly, the probability of being star forming is $1 - Q(r)$.

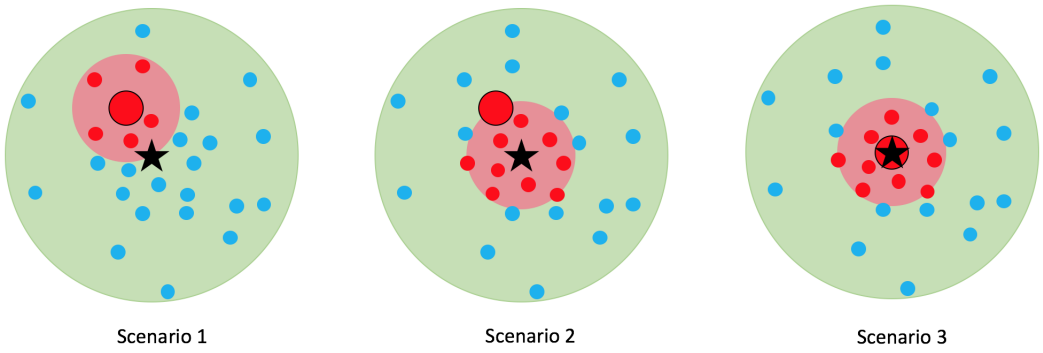


Figure 4.5: Illustration of the three possible ways for quenching to have spatial dependence within a halo. In each figure, the pale green symbolises the extent of the dark matter halo. The black star represents the centre of the halo. The large red circle represents a high mass galaxy, that may or may not be affecting the star formation rate of lower mass galaxies, denoted by the smaller dots. The shaded pale red area denotes 'quenched areas'. The lower mass galaxies are passive (denoted by being red) when in the quenched area, and star forming (denoted by being blue) when not. In Scenario 1, galaxies are preferentially quenched when near the massive galaxy, and the massive galaxy traces the halo profile. In Scenario 2, quenching is associated with proximity to the centre of the halo, and again all galaxies trace the dark matter profile. In Scenario 3, the massive galaxy is located at the centre of the halo, and quenching is associated with proximity to the centre of the halo, which coincides with the massive galaxy.

In conventional HOD modelling, galaxies are assumed to trace the profile of the halo, which I take to be a standard Navarro-Frenk-White (NFW, Navarro et al., 1996) profile.

The 1-halo term is then proportional to the convolution of the profiles each sample traces. If a galaxy is a 'central' then its 'profile' is taken to be a Dirac delta at the centre of the halo (this is the distinction between the central-satellite and satellite-satellite parts of the 1-halo term). I now discuss how to model each of the quenching scenarios within this framework:

Scenario 1 is the only scenario which is not equivalent to just modifying the profiles the samples trace - the other two are still symmetrical around the centre of the halo. The convolution of two NFW profiles is the probability of finding two galaxies at that separation. This pairing is then a massive-passive pair with probability $Q(r)$ and massive-SF with probability $1 - Q(r)$. Thus for the cross correlation of the massive environmental tracer population to low-mass passive galaxies:

$$1 + \xi_{1h}(\mathbf{r}) \propto Q(\mathbf{r}) \int_{\mathbb{R}^3} \rho_{\rm DM}(\mathbf{s}) \rho_{\rm DM}(\mathbf{r}-\mathbf{s}) d\mathbf{s}, \tag{4.3}$$

and for the cross correlation of massive to low mass star forming galaxies:

$$1 + \xi_{1h}(\mathbf{r}) \propto (1 - Q(\mathbf{r})) \int_{\mathbb{R}^3} \rho_{\rm DM}(\mathbf{s}) \rho_{\rm DM}(\mathbf{r}-\mathbf{s}) d\mathbf{s}. \tag{4.4}$$

Calculations involving NFW profiles such as these are commonly done in Fourier space, as an analytic expression for a NFW profile exists (Scoccimarro et al., 2001), and the convolution can be done as a multiplication. Note that this has a knock-on effect to the profiles the actual galaxies trace, with galaxies towards the centre of the halo preferentially being quenched (assuming that Q increases for small radii!). The passive low mass galaxies will then trace a profile

$$\rho_{\rm passive}(r) \propto \rho_{\rm DM}(r) \int_{\mathbb{R}^3} Q(\mathbf{s}) \rho_{\rm DM}(\mathbf{r}-\mathbf{s}) d\mathbf{s}, \tag{4.5}$$

and the star-forming low mass galaxies a profile

$$\rho_{\rm SF}(r) \propto \rho_{\rm DM}(r) \int_{\mathbb{R}^3} (1 - Q(\mathbf{s})) \rho_{\rm DM}(\mathbf{r}-\mathbf{s}) d\mathbf{s}. \tag{4.6}$$

This is due to the fact that the probability of the low mass galaxy being at a given point in the halo is proportional to the NFW profile. Then, once it is 'placed' there, the probability of being quenched is proportional to the integral over all space of the probability of being quenched, given all the possible places the massive galaxy can be.

Note that these expressions describe the global profiles of galaxies in halos of the given mass - individual halos will not necessarily be symmetric. Because of this complication, the auto-correlation is *not* just the convolution of this profile. To see why, imagine one dimensional halos in a one dimensional universe. Half the halos have two galaxies, each at a radius of 1 megaparsec away from the centre of the halo (e.g. at -1 and 1 if 0 is the centre of the halo). The other half of the halos have some sort of physical process that makes galaxies be at a radius of 0.5 megaparsecs (at -0.5 and 0.5). The true one-halo auto-correlation function of this universe would have peaks at 1 and 2 megaparsecs, the only

physical separations galaxies are found at. However the 'profile' has galaxies at -1,-0.5, 0.5 and 1. Taking the convolution of this would leave one to the mistaken conclusion that the auto-correlation function should have peaks at 0.5, 1, 1.5 and 2. This is another way of saying that averaging to find a profile and convolution do not commute. The correct way to find the auto correlation function in the 1D model is to find the auto correlation function for each halo, and then average, as opposed to averaging the halos, and then finding the auto correlation function.

Scenario 2 and Scenario 3 can be modelled by altering the profiles the lower mass galaxies lie on. Now they are just probabilistically passive or star forming depending on their radial location in the halo e.g. $\rho_{\rm passive}(r) \propto Q(r)\rho_{\rm DM}(r)$ and $\rho_{\rm SF}(r) \propto (1-Q(r))\rho_{\rm DM}(r)$. Then when the massive galaxy also traces the halo (Scenario 2), I obtain:

$$1 + \xi_{1h}(\mathbf{r}) \propto \int_{\mathbb{R}^3} Q(\mathbf{s})\rho_{\rm DM}(\mathbf{s})\rho_{\rm DM}(\mathbf{r}-\mathbf{s})d\mathbf{s}, \tag{4.7}$$

for the correlation with the passive low-mass population, and

$$1 + \xi_{1h}(\mathbf{r}) \propto \int_{\mathbb{R}^3} (1-Q(\mathbf{s}))\rho_{\rm DM}(\mathbf{s})\rho_{\rm DM}(\mathbf{r}-\mathbf{s})d\mathbf{s}, \tag{4.8}$$

for the correlation with the star forming low-mass population. When the massive galaxy is always found at the centre of the halo (Scenario 3), the second NFW profile in the convolution collapses into a Dirac delta, and the terms become:

$$1 + \xi_{1h}(\mathbf{r}) \propto Q(\mathbf{r})\rho_{\rm DM}(\mathbf{r}) \tag{4.9}$$

,

for the massive central to passive low-mass galaxies cross-correlation and

$$1 + \xi_{1h}(\mathbf{r}) \propto (1-Q(\mathbf{r}))\rho_{\rm DM}(\mathbf{r}).\text{,} \tag{4.10}$$

for the massive central to star forming low-mass galaxies cross-correlation. All three scenarios of course come with several assumptions. They all assume that the two low mass samples lie on 'complementary' distributions e.g. the sum of their profiles adds to an NFW.

—	Scenario 1	Scenario 2	Scenario 3
Massive Profile	$\rho_{\rm DM}$	$\rho_{\rm DM}$	δ
Passive Low Mass Profile	$\rho_{\rm DM} \times Q * \rho_{\rm DM}$	$Q\rho_{\rm DM}$	$Q\rho_{\rm DM}$
Star Forming Low Mass Profile	$\rho_{\rm DM} \times (1-Q) * \rho_{\rm DM}$	$(1-Q)\rho_{\rm DM}$	$(1-Q)\rho_{\rm DM}$
$1 + \xi_{1hP}$ (Cross)	$Q \times \rho_{\rm DM} * \rho_{\rm DM}$	$(Q\rho_{\rm DM}) * \rho_{\rm DM}$	$Q\rho_{\rm DM}$
$1 + \xi_{1hSF}$ (Cross)	$(1-Q) \times \rho_{\rm DM} * \rho_{\rm DM}$	$((1-Q)\rho_{\rm DM}) * \rho_{\rm DM}$	$(1-Q)\rho_{\rm DM}$
$1 + \xi_{1hP}$ (Auto)	NA	$(Q\rho_{\rm DM}) * (Q\rho_{\rm DM})$	$(Q\rho_{\rm DM}) * (Q\rho_{\rm DM})$
$1 + \xi_{1hSF}$ (Auto)	NA	$((1-Q)\rho_{\rm DM}) * ((1-Q)\rho_{\rm DM})$	$((1-Q)\rho_{\rm DM}) * ((1-Q)\rho_{\rm DM})$

Table 4.2: Summary of the resulting profiles, along with auto- and cross-correlation functions for different quenching mechanisms within a halo. δ denotes a Dirac delta, $*$ denotes a 3D convolution

Scenario 1 assumes that there is only one galaxy causing quenching within the halo, in practice there would be higher-order corrections for the small fraction of halos containing a higher number of quenching galaxies.

Note that in all three cases, the quenching mechanism gives a natural prediction for the fraction f of quenched low mass galaxies:

$$f = \frac{\int r^2 \rho_{\rm passive}(r) {\rm d}r}{\int r^2 \rho_{\rm passive}(r) {\rm d}r + \int r^2 \rho_{\rm SF}(r) {\rm d}r}. \tag{4.11}$$

This gives a fairly natural way to 'unify' the two models described in Simon et al. (2009), as the modified profiles dictate the expected number of galaxies in the halo when another galaxy sample is present. For example, the case where the occupation numbers are different, but the passive and active galaxies still trace the same profile, is described by $S = K$, and the relative amplitude between the two profiles changes, without changing the shape of their profile.

I summarise the profiles, auto-correlations and cross-correlations for each Scenario in table 4.2. Figure 4.6 shows the three scenarios for several example models. I use the same halo model for each plot, scale radius 1Mpc, and compactness parameter 10.

Note that necessarily if a profile or correlation function is larger on small scales it must be smaller on large scales, as they are normalised to have total weight of unity. All the panels in Fig. 4.6 are normalised, and the cross-over sometimes occurs out of the range plotted. However typically the amount by which the flatter distribution is greater than the more compact one on large scales is tiny, as it adds up to a greater probability due to it being weighted by r^2 when integrated over the whole volume.

The auto-correlation functions for Scenario 1 as discussed above do not present a straightforwards analytic expression. However, it is still possible to calculate, by a Monte-

Carlo method. I sample a location within the halo for the massive galaxy to be. I then sample two locations for the lower-mass galaxies to be, and sample whether they are passive or star forming. When a pair of the correct type is found, their spatial separation is recorded, which essentially allows DD to be constructed. The auto-correlations for Scenario 1 in Figure 4.6 are constructed in this manner.

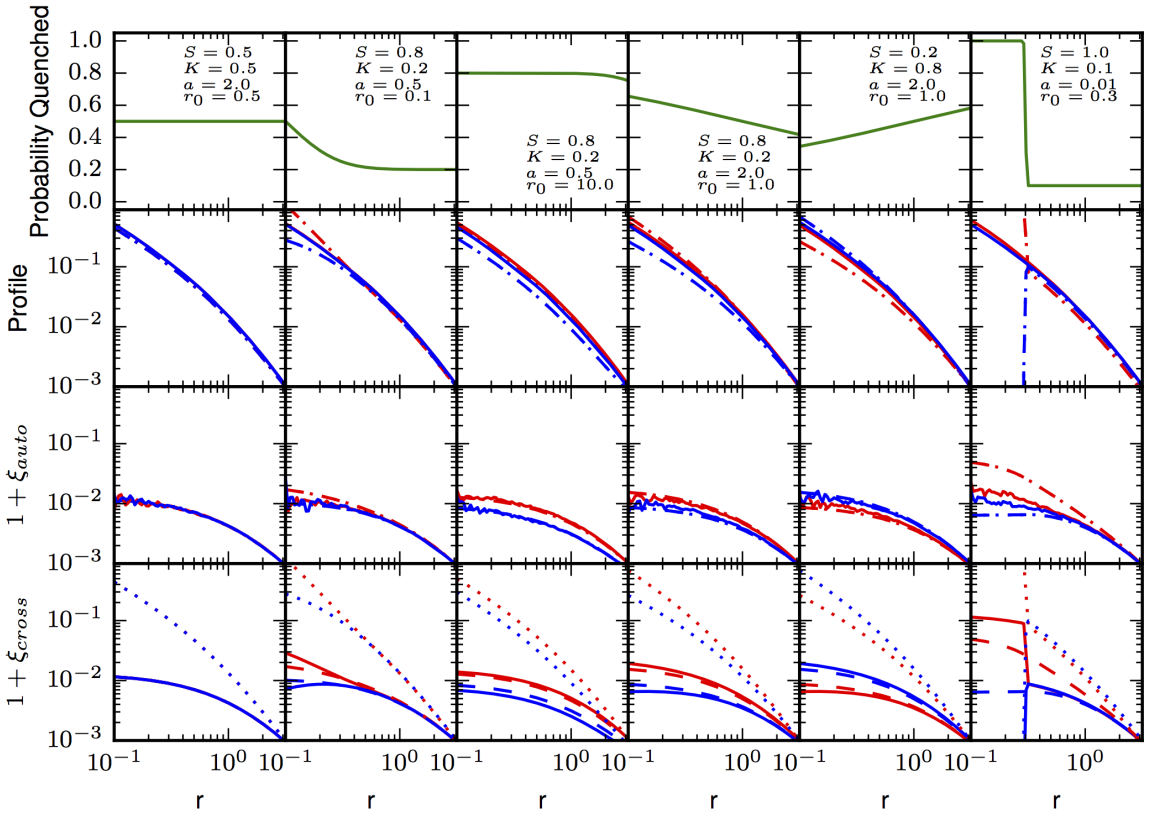


Figure 4.6: An illustration of the consequences of different quenching models within my formalism. First row: probability of being quenched as a function of spatial separation. Second row: profile of galaxies within a halo. Third row: auto correlation of galaxy samples. Fourth row: cross correlation of low mass to high mass galaxy samples. Full lines correspond to Scenario 1, dashed lines to Scenario 2, dotted lines to Scenario 3. All functions (described in table 4.2) normalised, red corresponds to passive and blue to star forming where appropriate. The differences between the Scenarios are clearly much more prominent in the cross-correlations than the auto-correlations for all models.

In the first quenching model in Figure 4.6, there is no radial dependence for quenching. The profiles the galaxies trace are NFW, the auto-correlation function is its self-convolution, and the cross correlations are the NFW self-convolution in Scenario 1 and 2, and the con-

volution of NFW and a point for Scenario 3 (the standard central-satellite term.

In the second quenching model, there is a slight increased tendency for galaxies to be quenched at small separations (either from each other, or from the centre of the halo depending on the scenario). It can clearly been seen that on small scales passive galaxies become more concentrated at the centre of the halo. This has corresponding effects on the auto and cross-correlations, both being enhanced/depressed on small scales. The third quenching model shows the same, but the characteristic length scale for quenching is a lot larger, and the enhancement can be seen to hold on larger scales.

The fourth quenching model shows a perhaps more realistic model, giving measurables closer to the data. In particular the Scenarios have different impacts on the differences between the cross-correlations, but have near-identical impacts on the auto-correlations. This shows potentially how they could be distinguished. The fifth quenching model is identical to the fourth, except swapped, in the sense that galaxy star formation is enhanced when close to other galaxies/the centre of the halo, and, as expected, the inverted result is obtained.

The sixth model shows an extreme case, where galaxies are quenched with certainty at separations closer than the characteristic quenching scale. This causes the profile of the star forming galaxies to go to zero below a certain radius in Scenario 2 and 3, as no galaxies can be star forming in the centre. The auto correlations are enhanced/depressed to different degrees, as normal. But the cross correlations show the most interesting case. For Scenario 1, the cross correlation for star forming galaxies goes to zero, as massive galaxies are never seen near to star forming galaxies. Similarly for Scenario 3, but in Scenario 2, there is still some cross-correlation, as it is still possible to find massive galaxies near star forming galaxies when they are both away from the centre of the halo.

This illustrates that, within the context of this model, judicious measurements of auto and cross correlation functions can allow one to distinguish between different quenching powers, models and scenarios.

What is the physical interpretation of the three scenarios? In Scenario 1 quenching is dependent on physical proximity between the galaxies. Detection of quenching within a halo being geometrically distributed in this manner could reasonably be interpreted as

supportive of quenching mechanisms like splashback, tidal stripping, harrassment, an AGN heating up its vicinity, and so on. Conversely, for Scenario 2, quenching is dependent on location within halo, and the correlation functions being supportive of Scenario 2, could be interpreted as quenching mechanisms like ambient temperature in the halo heating the gas in the galaxy. Scenario 3 is the case when the two coincide geometrically because the massive galaxy is at the centre of the halo, which makes it harder to distinguish mechanisms in the simple version of the model discussed here.

The reason in this work I do not also fit for the autocorrelations is that there will be contributions from other halo masses, which complicates matters because a) the quenching mechanism is likely halo and stellar mass dependent b) the number of quenching sources is no longer fixed at one. However I propose that the future direction of inquiry attempts to model the cross and auto correlation functions simultaneously, which I discuss further in section 4.5.3.

The language used in this paper is motivated by results such as Woo et al. (2013), who suggest that quenching is associated with location in a halo, or proximity to another galaxy. However I note that it may be possible to get similar geometric arrangements of galaxies without there being a physical interaction causing this effect - perhaps passive satellite galaxies were in more massive halos before they were accreted, and that when these halos become sub halos they are more likely to be concentrated towards the centre of the larger halo by some mechanism of the sub-halo dynamics. Furthermore, the "characteristic quenching scale" is purely a 'measured' dependence, which is not necessarily the same as the physics of the actual mechanism because the motion of the satellites within the halo could 'smear' the signal e.g. suppose an AGN quenches star formation in every galaxy within ~ 0.1 Mpc, (e.g. Rawlings & Jarvis, 2004), galaxies will be observed to be quenched at larger radii than that because their orbit might take them close to the AGN, permitting them to be quenched, but then move them further away again.

Note that this formalism still strictly describes quenching and conformity within a halo. Hearin et al. (2016a) suggest their decorated HOD model may be able to explain 2-halo conformity though. In addition, the quenching parameters are strictly dependent on the sSFR cut to divide into star forming and passive, and will in general be different for different

sSFR and stellar mass cuts. In a scenario where galaxies were passive at the centre of a halo, and got progressively more star forming towards the outside, one might expect r_Q to increase as one increased the sSFR cut.

4.4.3.2 The Halo Model

Halos are non-linear small scale overdensities in the dark matter density field caused by collapse when an overdensity reaches a critical amount - in this sub-section I discuss the model and parametrisation of halo profiles that I use in the model. Both theoretical considerations and N-body simulations suggest that NFW profiles (Navarro et al., 1996) describe halo density profiles well. NFW profiles are parametrised by a scale radius R_s which denotes scale of transition from the inner r^{-1} density profile and the outer r^{-3} density profile, a virial radius within which the dark matter has virialised, a compactness parameter which denotes the radius between the two, and a characteristic density which links to the halo mass. As the untruncated NFW profile has infinite total mass (is unnormalisable), halo models typically assume that galaxies only lie within the virial radius and that the halo density drops to zero at this point. Note that it is necessary to make some sort of similar assumption, otherwise the probability of finding galaxies within some radius is proportional to the mass enclosed within this radius, which is finite compared with the infinite mass as the radius extends to infinity. The total mass of the halo is then the density integrated over the sphere with radius R_{vir}.

There are multiple different prescriptions describing how halos of different masses and compactnesses link to overdensities in the global dark matter field (e.g. see Coe, 2010, for an overview). I use the parametrisation described in Coupon et al. (2012). In this formalism, the critical overdensity $\Delta(z)$ required for virialisation is taken from Weinberg & Kamionkowski (2003), the relationship between the scale radius and the virial radius is given by $R_{vir} = c \times R_s$, with c a function of halo mass and redshift as given in Coupon et al. (2012).

4.4.3.3 Fitting Quenching Models

I seek to model the environmental role of quenching with the observations, using the results for the 1-halo term found in section 4.4.3.1. The theory as developed there only describes the 1-halo term resulting from galaxy interactions for a single halo mass. To fully model the cross-correlation function this must be generalised (as in conventional HOD) to include all halo masses, which is likely not straightforward as the quenching power is likely to be a function of halo mass, and the mass of the galaxy it is acting on. However the massive galaxies live in massive halos, where the HMF is dropping off quickly, which allows us to approximate that the 1-halo term is dominated by galaxies in a small range of halo masses.

Models based solely on Scenario 1 and 2 produced extremely poor fits and were in general unable to reproduce my results, suggesting the massive galaxy must typically be at the centre of the halo (one can see by eye in figure 4.6 that if there is no central galaxy the one-halo term is very shallow at small radii). This is consistent with the results in Chapter 3 that ~ 90 per cent of galaxies of these stellar masses at these redshifts are central galaxies. The general expression for the one-halo central-satellite term is

$$1 + \xi(r) = \int_{M_{vir}}^{\infty} n(M) \frac{\langle N_{cen} \rangle \langle N_{sat} \rangle}{n_1 n_2} \rho(r, M) \mathrm{dM} \tag{4.12}$$

(see Coupon et al., 2015, where I remove the factor of 2, as it is cross correlations under consideration and it is not nescessary to account for each pair being counted twice), where $n(M)$ is the normalised HMF in counts per cubic comoving megaparsec per unit mass, ρ is the normalised density of the profile per cubic comoving megaparsec, n_i is the comoving number density of each type of galaxy, $\langle N_{cen} \rangle$ is the expected number of centrals, $\langle N_{sat} \rangle$ the expected number of satellites.

Because the satellite number rises as a power law, and the HMF drops off as an exponential, the integral is dominated by a relatively small range of halo masses, so the profile dependence on halo mass can be taken outside the integral, and the complication of halo mass dependence on quenching power ignored. This leaves:

$$1 + \xi(r) \approx \int_{M_{vir}}^{\infty} n(M) \frac{\langle N_{cen} \rangle \langle N_{sat} \rangle}{n_1 n_2} \mathrm{dM} \times \rho(r, M_{dom}), \tag{4.13}$$

where M_{dom} is the dominant halo mass scale probed/contributing to the integral. I take $\rho(r)$ to be a normalised NFW, modified as per Scenario 3, and calculate the integral assuming the halo occupation measured in Chapter 3, modifying the n_i appropriately. In particular I take $\langle N_{sat} \rangle = f \times N_{HOD}$, where f is the implied fraction of satellite galaxies quenched according to equation 4.11 and N_{HOD} is the number of satellites in a halo of that mass implied by the HOD fits in Chapter 3. If I were doing a simultaneous fit of both occupation numbers *and* a quenching mechanism, I could take n_i and N_{HOD} to be properties of the model, but this ansatz will suffice for the purposes of this chapter.

Equation 4.12 shows amplitudes of the profiles depend on the passive fraction of the population as a whole, and the passive fraction of the galaxies in the vicinity of the massive galaxy. In general for my measurements, the overall amplitude is set by the halo mass the interactions are taking place in, and the ratio R between the two cross-correlations in the 1-halo term is set by:

$$R = \frac{f_E^P / f_G^P}{f_E^{SF} / f_G^{SF}} = \frac{f_E^P / f_G^P}{(1 - f_E^P)/(1 - f_G^P)}, \tag{4.14}$$

where f_E^P is the passive fraction of the low mass galaxies in the environment traced by the massive galaxies, f_G^P the same fraction globally (e.g. for all low mass galaxies), and f_E^{SF} and f_G^{SF} the corresponding star forming fractions.

Once the one-halo cross-correlation term is constructed, I then add the two-halo term. I then project into angular space using the source redshift distribution and subtract the integral constraint.

In total this gives seven parameters to describe each individual set of measurements: two large scale cross biases, four parameters to describe the galaxy interactions, and one parameter for halo mass. I use EMCEE (Foreman-Mackey et al., 2013) to provide a Markov chain Monte Carlo sampling of the parameter space to fit the correlation functions. I use a uniform prior over $0.5 < b_{\times 1}^2 < 9$, $0.5 < b_{\times 2}^2 < 9$, $11 < \log_{10}(M_{\rm h}/M_\odot) < 15$, $-2 < \log_{10} r_Q < 0$, $0 < S < 1$, $0 < K < 1$ and $0 < a < 3$ (that is to say uniform in log space for mass and squared space for the biases). I used 20 walkers with 1000 steps, which have starting positions drawn uniformly from the prior.

The likelihood is calculated using χ^2 from both clustering measurements:

$$\chi^2 = \sum_i \frac{[\omega_{\rm pas}^{\rm obs}(\theta_i) - \omega_{\rm pas}^{\rm model}(\theta_i)]^2}{\sigma^2_{w_{i_{\rm pas}}}} + \sum_i \frac{[\omega_{\rm SF}^{\rm obs}(\theta_i) - \omega_{\rm SF}^{\rm model}(\theta_i)]^2}{\sigma^2_{w_{i_{\rm SF}}}}, \tag{4.15}$$

where $\omega_{\rm pas}^{\rm obs}$ is the observed massive to passive cross correlation function data, $\omega_{\rm pas}^{\rm model}$ is the corresponding model correlation function, $\sigma_{w_{i_{\rm pas}}}$ is the error on these measurements, θ_i are the angular scales fitted for, separated by more than the smoothing length used to estimate the correlation function (with SF corresponding to star forming for the other variables).

4.4.3.4 Results of the quenching model

I show the data and corresponding fits for the four redshift bins in Figure 4.7, and the triangle plot for the MCMC run of the $1 < z < 1.5$ redshift bin in Figure 4.8 to highlight the typical posterior probabilities for each parameter. Good fits are clearly obtained, capturing the key features of the data. Figure 4.9 shows the corresponding implied probabilities of being quenched as a function of radius within the halo. The dotted lines show the global probability (the fraction of all galaxies that are passive, see table 4.1) of the low mass galaxies being quenched at that redshift.

At the highest redshift, the galaxies are no more likely to be quenched when inside the environment associated with the massive galaxy, than out of it. As noted in section 4.2, the effects of dust may mean the true passive fraction may be slightly lower, but the fact that the cross-correlation functions overlap means that the global passive fraction and the passive fraction in the massive galaxy halo must be similar. At intermediate redshifts $1 < z < 1.5$ there is moderate radial dependence within the halo for the likelihood of being quenched - on the outskirts of the halo the probability of being quenched is ~ 0.55 (similar to the global passive fraction at this redshift), but this rises to ~ 0.85 in the centre of the halo. However by the lowest two redshift bins($0.4 < z < 1.2$), the lower mass galaxies are preferentially quenched in halos hosting massive galaxies (a passive fraction of ~ 0.85 within the halo, versus ~ 0.65 outside), with very little radial dependence within the halo.

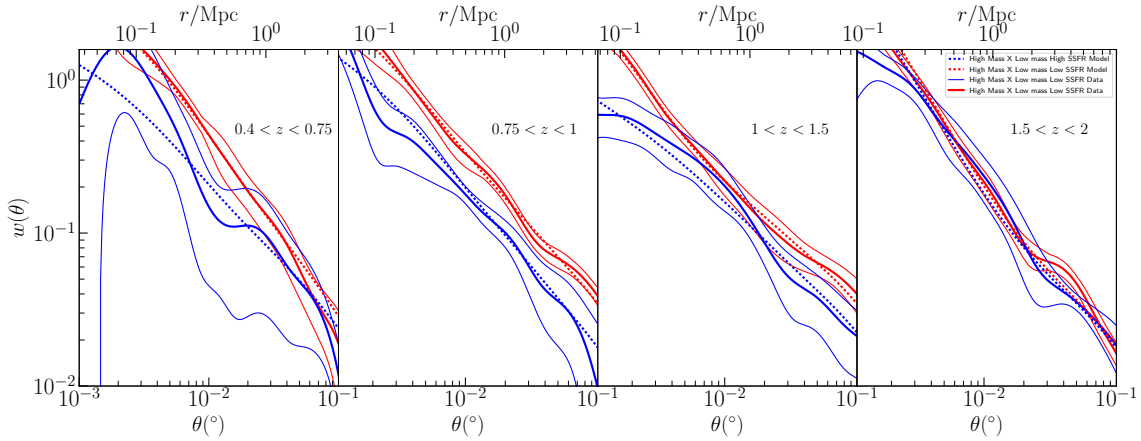


Figure 4.7: The 'environment tracer' measurements and best fit model at my four redshift ranges. The solid lines are the data; the cross correlations of massive to less massive galaxies, red when the low mass galaxy is passive and blue when star forming. The error bars are the 16th and 84th percentiles of the estimator. The dashed lines correspond to the best fit model model using the quenching modification to the 1-halo term.

I show in Fig. 4.10 the halo mass of the best fitting profile for each redshift bin, which can be seen to increase with time. This mass should not be directly interpreted as the halo mass that all the galaxies are in (typically the galaxies are found in a range of halo masses), but as the mass corresponding to the scale at which non-linear effects are needed to describe the cross-correlation functions. I interpret this increase as resulting from evolution in the HMF; at high redshift, the more massive halos are rare, and lower mass group environments are being predominantly probed as they are more common. At lower redshifts, the higher mass halos aren't as rare, and as the more massive halos have more satellites in them, they become more dominant in my measurements, meaning that non-linear effects extend to larger radii.

Figure 4.11 shows the cross biases of my measurements. Note that these are not the biases of each sample of galaxy, they are the cross biases, which is the geometric mean of the two samples being cross-correlated. Typically one would expect that at a given redshift the higher mass sample would have a higher bias than the lower mass sample, so the cross bias will be between the two values. In each redshift bin the masses of the two samples are unchanged, it is just the star formation rate of the lower mass sample that is changed. I see a general trend of bias increasing at high redshifts, from $b \sim 1$ at $z \sim 0.6$ to $b \sim 2.5$

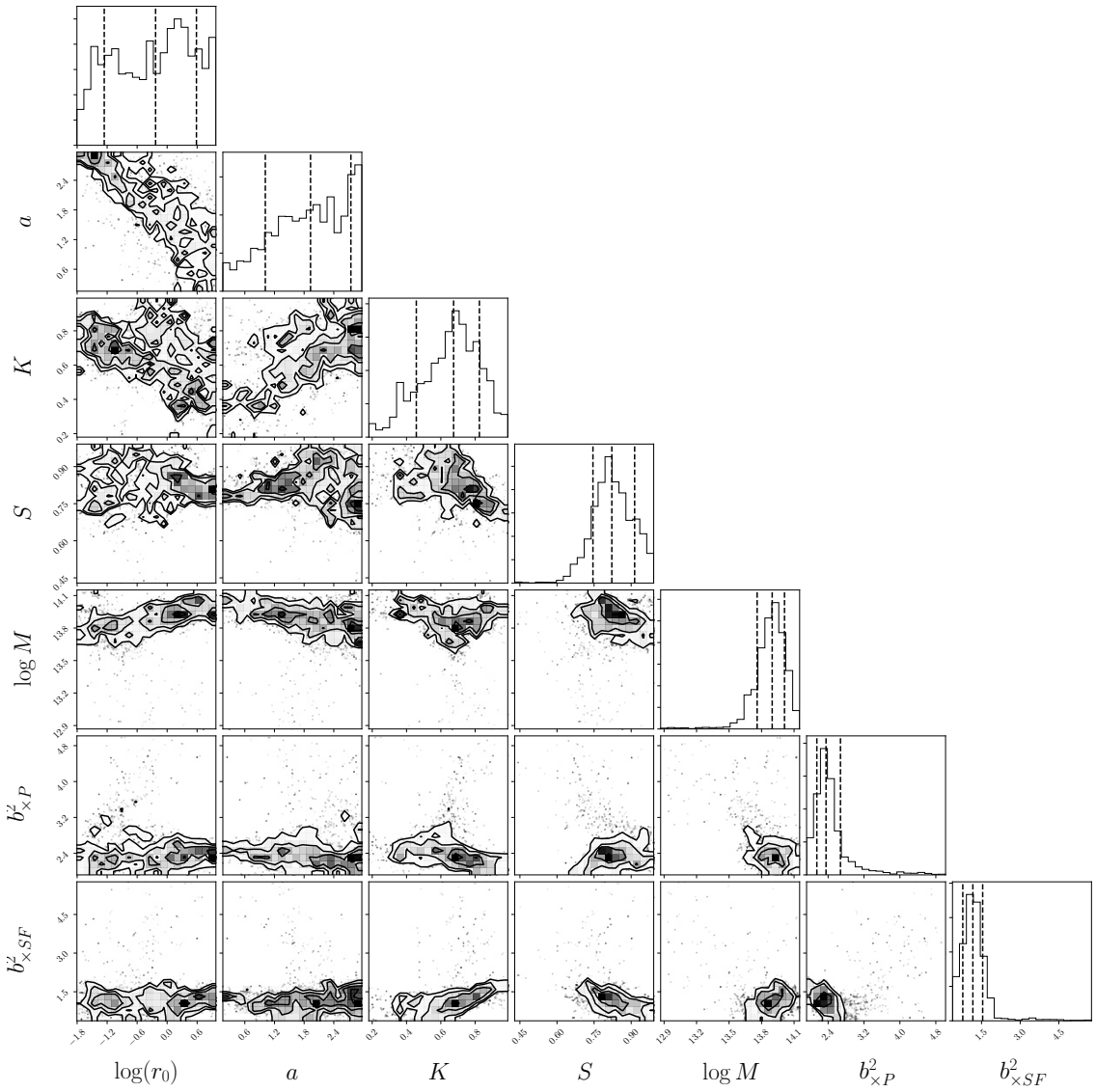


Figure 4.8: Corner plot of the posterior for the MCMC parameter estimation for the $0.75 < z < 1$ quenching model.

at $z \sim 2$, which is consistent with measurements in McCracken et al. (2015) and Chapter 3. However I see that in the highest bin, the biases are essentially the same regardless of the star formation rate of the lower mass sample, but diverge for lower redshifts, consistent with what I qualitatively described in section 4.4.1.

In terms of length scale of quenching r_Q and sharpness of transition a, little/no constraints can be made, as typically little radial dependence for quenching probability is found (apart from the $1 < z < 1.5$ bin). In table 4.3 one can see that typically values of around 2

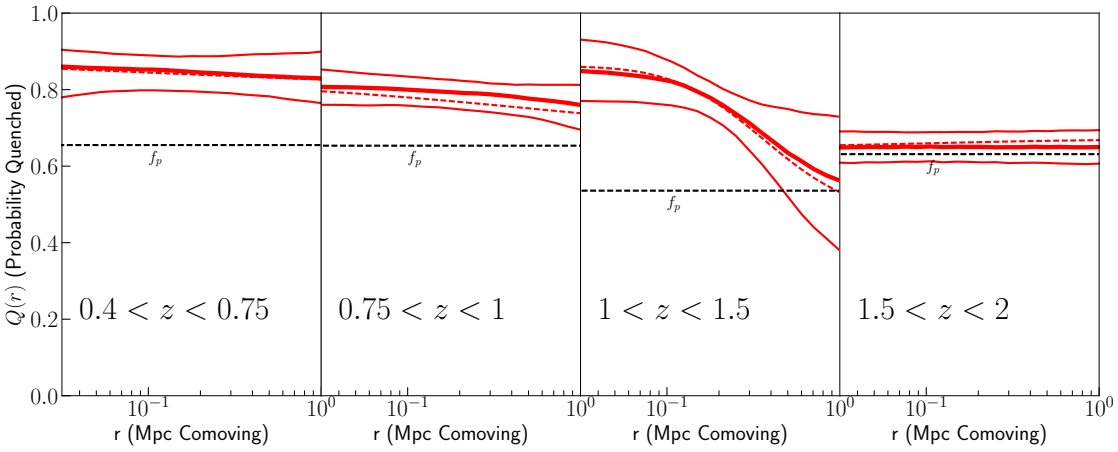


Figure 4.9: The quenching probabilities implied by the MCMC parameter fitting for the four redshift ranges is shown in red, i.e. the probability of the lower mass galaxies being quenched as a function of radial location in halo. The full thick and two thin lines are the pointwise median and 16th and 84th percentiles respectively of the sampled models. The dashed red lines are the model corresponding to the best fit parameters. The horizontal black dashed lines are global passive fraction for the lower mass galaxies at that redshift.

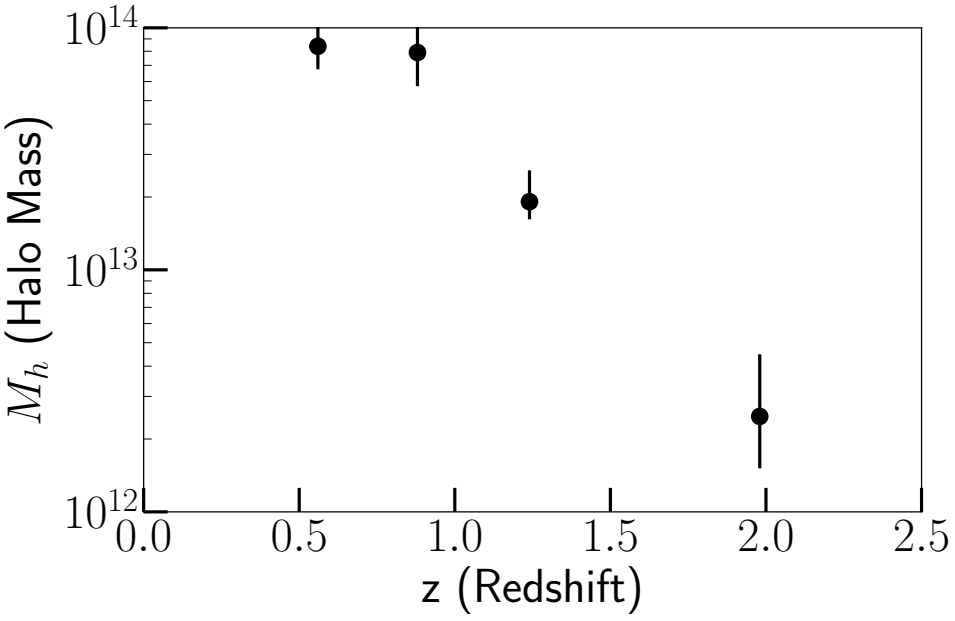


Figure 4.10: The halo mass of the environment probed at each redshift I make measurements for. Error bars are the 16th and 84th percentiles of the MCMC fitting samples.

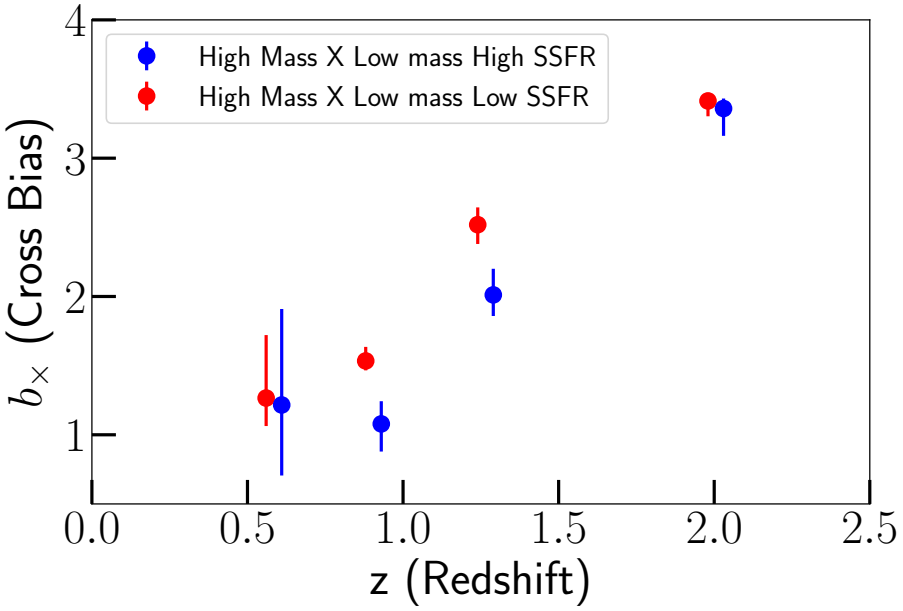


Figure 4.11: The cross biases of each of my correlation functions. The cross bias is the geometric mean of the bias of each subsample used in the cross-correlation.

Table 4.3: The constraints on our quenching and bias model from our MCMC results. The errors are the 16th and 84th percentiles of the posterior. The cross biases are labeled by what pairs of samples they correspond to; H-PL corresponds to the high mass sample to low mass passive sample, H-SL the high mass to low mass star forming sample.

Redshift Range	K	S	r_Q	a	$b_\times$ (H-PL)	$b_\times$ (H-SL)	M_h
$0.40 < z < 0.75$	$0.808^{+0.13}_{-0.16}$	$0.885^{+0.084}_{-0.3}$	$0.083^{+0.52}_{-0.064}$	$2.23^{+0.47}_{-0.77}$	$1.27^{+0.46}_{-0.2}$	$1.22^{+0.69}_{-0.51}$	$13.9^{+0.3}_{-0.095}$
$0.75 < z < 1.00$	$0.674^{+0.15}_{-0.22}$	$0.822^{+0.092}_{-0.076}$	$0.588^{+3.3}_{-0.53}$	$1.93^{+0.82}_{-0.92}$	$1.53^{+0.1}_{-0.068}$	$1.08^{+0.16}_{-0.2}$	$13.9^{+0.11}_{-0.14}$
$1.00 < z < 1.50$	$0.478^{+0.23}_{-0.29}$	$0.861^{+0.086}_{-0.094}$	$0.362^{+0.59}_{-0.22}$	$0.576^{+0.55}_{-0.26}$	$2.52^{+0.13}_{-0.14}$	$2.01^{+0.19}_{-0.15}$	$13.3^{+0.13}_{-0.073}$
$1.50 < z < 2.00$	$0.671^{+0.08}_{-0.11}$	$0.673^{+0.14}_{-0.09}$	$0.109^{+0.25}_{-0.061}$	$1.07^{+0.69}_{-0.56}$	$3.41^{+0.037}_{-0.11}$	$3.36^{+0.072}_{-0.2}$	$12.4^{+0.26}_{-0.21}$

are found for a, meaning that changes in quenching probabilities span ~ 2 orders of magnitude i.e. no sharp transitions - again except for the $1 < z < 1.5$ bin, which has $a \sim 0.5$, for a slightly sharper transition.

I also model the results of section 4.4.2 when splitting by sSFR in the massive galaxy sample, see figure 4.12. It is evident that lower mass galaxies are preferentially quenched when the massive galaxy is passive as opposed to star forming.

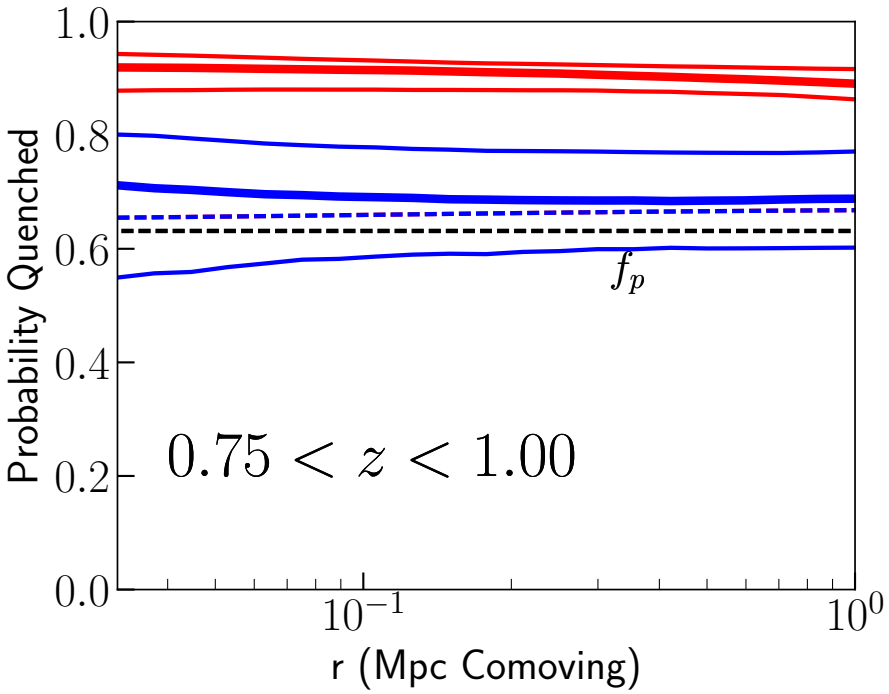


Figure 4.12: Estimate of the probability of being quenched as a function of radial location from my best fit models when the central is passive (red), and when the central is star forming (blue). Full lines correspond to the pointwise 16th, 50th and 84th percentiles, dashed line to the quenching model of the best fitting parameters.

4.5 Discussion

4.5.1 Summary of Results

My results in sections 4.4.1 and 4.4.2 show that targeted cross-correlation measurements (the 'environment tracer' method) can be used to probe the environmental quenching and galactic conformity noted in other measurements. Modelling this within the HOD framework shows that I am probing group environments of $\sim 10^{13} M_{\odot}$ halos, which have a massive galaxy at the centre, and 1-2 satellites of lower masses, which agrees with Viola et al. (2015).

The measurements made over a range of redshifts suggest a coherent 'timeline' of the role of environment/the colour-density relation over the age of the Universe. The data suggests that environmental quenching is not a significant factor for the galaxies under consideration in this study at $z \sim 2$, if it was there would be differences in the clustering that depended on the star formation rates of the galaxies. In the $1 < z < 1.5$ bin, environmental quenching

has started to play a role on scales smaller than $\sim$ 1Mpc, and by $z \sim 0.5$ environmental quenching has played such a part that passive galaxies are found in dramatically different environments to star forming galaxies of the same mass and have different biases etc.

I included in my model the probability of a galaxy being quenched having a radial dependence within the halo; I found that the data favoured a rising probability of quenching at intermediate redshifts ($1 < z < 1.5$), and a flat probability at lower redshifts ($z < 1$). There are several possibilities for why this may be. It could be that the motions of the satellites within the halo quickly smear out any signal of spatial dependence of quenching e.g. satellites are preferentially quenched in the centre of halos, but then quickly move out again, weakening the signal. Alternatively it could be possible that quenching is associated with time spent in a halo; Smethurst et al. (2015) find that galaxies closer to the centre of a group halo were quenched earlier. This time dependence, combined with the fact that sub-halos accreted earlier are typically found towards the centre of the halo[1], could mean that at $z \sim 0$ a radial dependence is observed as there is a large range of accretion times of sub-halos, but at the comparatively early times probed in this study there was a much smaller range of accretion times for sub-halos, limiting the possible radial dependence.

My results show that the satellite quenched fraction is lower when the massive galaxy is star forming (although still higher than the global fraction). I interpret this as a measurement of how 1-halo conformity can manifest itself in the correlation function. Galaxies are preferentially quenched as satellites when the central galaxy is quenched, and vice versa. I also see 2-halo conformity in my results in the sense that on scales larger than individual halos, the difference between the passive/star-forming low-mass cross-correlation amplitudes is greater when the massive galaxy is passive than when star-forming. Note I do not attempt to distinguish between the weaker sense of 2-halo conformity (correlations that are a natural result of the fact that massive halos are typically near other massive halos, and massive objects typically host more quenched objects) and the stronger claim that there is an environmental effect over volumes of space much larger than halos that can have an effect on galaxies. Hearin et al. (2016a) note that 1-halo conformity and 2-halo conformity

[1]It appears to be the case that sub-haloes are distributed independently of sub-halo mass, but with some accretion time dependence, within the larger halo, but see Han et al. (2016) for a more detailed discussion.

are not necessarily separate effects, as satellites lived most of their lives as nearby central galaxies before accretion. My results are essentially consistent with the view presented using SDSS data at $z \sim 0$ (Knobel et al., 2015; Paranjape et al., 2015; Pahwa & Paranjape, 2016) that conformity is strong within haloes, but that correlations on scales larger than halo scales are simply a result of halo bias. Knobel et al. (2015) use the groups and halo masses of Yang et al. (2012), and find that satellite galaxy sSFR is correlated with the sSFR of its central galaxy, at fixed halo mass, central and satellite stellar mass, distance from the centre of halo, and large-scale overdensity. Paranjape et al. (2015) show by generating mock catalogues that halo-halo correlations are not required to generate a signal on scales larger than individual haloes. Pahwa & Paranjape (2016), similarly to this chapter, develop an approach to modelling the effects of conformity on clustering measurements using a modification of HOD schemes. They find SDSS clustering measurements suggest similar levels of 1-halo conformity to Knobel et al. (2015), with no evidence for halo-halo correlations. Zu & Mandelbaum (2015, 2016) also provide a highly relevant evolution of the HOD model, fitting to clustering and galaxy-galaxy lensing in SDSS, finding that models with halo mass as the sole driver of probability of being quenched fit the data best. Tinker et al. (2017) also find that halo-halo correlations are negligible using halo age matching models. At higher redshift, Kawinwanichakij et al. (2016) investigate conformity at $0.3 < z < 2.3$, finding that conformity is most significant at to $0.6 < z < 1.6$, and Berti et al. (2017) investigate $0.2 < z < 1$, finding strong evidence for substantial conformity on scales < 1Mpc, and marginal evidence for a small amount of conformity on scales $1 - 3$Mpc. There does certainly seem to be some dependence on the large-scale density field in some contexts (e.g. the stellar to halo mass ratio is found to have a slight dependence on it, Cen & Safarzadeh, 2014), but in summary most of the literature, and the work presented here, seems to find that conformity is only relevant on the scales of individual haloes, but can be significant up to $z \sim 1.5$.

4.5.2 Quenching over cosmic time

Direct comparisons of measurements like 'passive fraction' and 'red fraction' in the literature are difficult a) because these measurements are functions of population, so comparisons are

only direct between matched stellar mass ranges and redshifts etc. b) these fractions are dependent on choice of dividing colour or sSFR break and c) photometric sSFR values remain tentative, and can vary with what photometry is available and choice of templates etc. Nonetheless it is possible to see whether the broad picture suggested by my results is consistent with other studies.

Observationally, Woo et al. (2013) used the SDSS ($z \sim 0.1$) to investigate the role of halo mass and location in the halo on quenching. Their key findings were that environmental effects on quenching seemed more correlated with halo mass at fixed stellar mass, than with stellar mass at fixed halo mass, suggesting that halo mass was the more important factor. They found that quenched fraction increased with halo mass, and increased towards the centre of the halo, most notably in the central $0.1 - 0.3R_{vir}$. This appears to be in contradiction to my results, which exhibit little radial dependence at lower redshift, but closer inspection shows that at $\sim 10^{13} M_{\odot}$, the halo mass being predominantly probed here, they find the quenched fraction is flat at about 0.5 for the inner part of the halo, and drops off to 0.2 at around $0.6R_{vir}$, so future work is needed to see if clustering recovers the radial dependence they measure. In the GAMA survey, Prescott et al. (2011) find similar results of quenched satellite fraction increasing with the mass of the central galaxy, with little radial dependence at lower masses, and some radial dependence on small scales for the more massive galaxies. In agreement with my results/1-halo conformity, they also find that the quenched fraction is higher when the central is passive (see their figure 7).

In terms of simulations, Gabor & Dave (2014) discuss a cosmological simulation with a prescription for environmental quenching within halos in areas dominated by hot gas. A key finding is that a lot of what might be interpreted as 'environment' quenching is really 'mass quenching' in 'pre-processed' satellites; satellites at $z = 0$ used to be centrals at higher redshift and experienced mass quenching before accreting and becoming a satellite. In addition they find that, although with larger scatter, galaxies closer to the centre accreted earlier i.e. galaxies towards the edge of the halo have been there for a shorter period of time, suggesting there is also a radial component to how long galaxies have been in the environment. Similarly a fraction of galaxies are 'ejected' so it is possible to have galaxies that get environmentally quenched, but then are found far away from that environment.

However they find that there is a strong dependence on the quenched fraction on both radial location in halo, and halo mass (with their results agreeing with Woo et al., 2013). They find a quenched fraction of ~ 0.3 everywhere in the halo for halo masses $\sim 10^{12} M_{\odot}$, spanning 0.3 at the edge of the halo to near unity at the centre of the halo for halo masses $\sim 10^{13} M_{\odot}$, and nearly all quenched except for the very edge of the halo for $> 10^{13} M_{\odot}$. This appears to have slightly more radial dependence than I find in my results, but is at $z = 0$, so isn't directly comparable to my measurements. In addition, at the halo masses comparable to my results, the quenching probability is actually quite flat apart from a) the very centre of the halo ($< 0.05 R_{vir}$ and b) beyond the virial radius, which I can't account for within this formalism, so it is not clear if my results of little radial dependence are in tension or not.

Both Woo et al. (2013) and Gabor & Dave (2014) are at $z \sim 0$. In terms of $z \sim 1$, there is substantial disagreement - some authors claim that the sSFR-density relation is qualitatively the same as $z \sim 0$ relation, while other authors have suggested that the sSFR-density relation actually essentially swaps at this epoch e.g. sSFRs are higher in denser environments, in contrast to the local universe. Patel et al. (2009) and Baldry et al. (2006) both report the $z \sim 0$ relation holding at $z \sim 1$ in Spitzer and Gemini data respectively. Nantais et al. (2017) find strong evolution in the quenched fraction of galaxies in clusters - that at $z \sim 1.6$ the quenched fraction is similar in clusters to in the field, but by $z \sim 0.9$ galaxies have a much higher quenched fraction in clusters than in less dense environments. Duivenvoorden et al. (2016) find in the Herschel Extragalactic Legacy Survey (HELP) that their data is consistent with no environmental dependence at $z > 2$. Conversely, Elbaz et al. (2007) and Welikala et al. (2015), find the relation reversing in the Great Observatories Origins Deep Survey and South Pole Telescope data (with some simulations reporting similar results e.g. Tonnesen & Cen, 2014). My results are in general not supportive of the relation reversing (although more work is needed to confirm that environments and host galaxies etc. are being compared consistently) - the picture presented by the data is that of the relation weakening, until it only holds on small scales by $z \sim 1.2$, and disappearing by $z \sim 2$. If the relation reversed, by higher redshifts one might expect to see the star forming lower mass galaxies being more clustered around

the tracers of dense environments, the massive galaxies, and one might expect the biases of star forming galaxies to be higher than that of passive galaxies of the same mass at higher redshifts, which I do not see. However it will take more work on the development of incorporating quenching mechanisms into the HOD/clustering framework before I can be sure I am consistently comparing the same galaxies in the same environments between different probes of density.

With regards to understanding why it might be that the sSFR-density relation emerges only at later times, Peng et al. (2010) offers some possible insight. They study environment in SDSS and zCOSMOS up to $z \sim 1$, looking at red/blueness in relation to local overdensity δ, and explicitly attempt to empirically model the measurements without the invocation of physical models e.g. the idea that galaxies live and are formed in halos. They find, among other things, that the differential effect of a given environment doesn't change with epoch, but that an environmental dependence of f_{red} emerges over time, because galaxies move to higher overdensities as large scale structure grows in the Universe, not because there is any change over time of the effect of a given environment. They have a relative quenching efficiency, which characterises at fixed stellar mass how much more quenched galaxies are in some environment compared with some reference environment. This is in broad agreement with the picture suggested by my results, of the environment not yet having had enough time to have had effect at high redshift, and only coming into significance later than $z \sim 1$. For understanding why there appears to be little radial dependence to the probability of quenching, McGee et al. (2014a) discuss an updated view of the baryon cycle and the quenching of accreted satellite galaxies in halos. One paradigm of satellite quenching is that galaxies that are accreted are forming stars from a reservoir of gas, that is slowly removed from the stellar disk in the case of ram pressure stripping, or from the host sub-halo in the case of strangulation. However McGee et al. (2014a) discuss observations and simulations that suggest a slightly different view - that star formation in field galaxies is driven by fresh inflows, as opposed to feeding on a reservoir. In this scenario, quenching of galaxies when they are accreted by a larger halo is a result of these inflows stopping, as opposed to gas from a reservoir being removed by any sort of stripping. McGee et al. (2014a) discuss evidence that these inflows are stopped when the accreted galaxy passes

$\sim 1.8R_{vir}$, which could explain the lack of radial dependence; quenching merely depends on passing that radius, as opposed to some radius dependent stripping process.

With regards to the use of correlation functions to make inferences about environmental effects, Skibba et al. (2009) use *marked* correlation function statistics using SDSS Galaxy Zoo data, where essentially the conventional correlation function is compared with a weighted correlation function e.g. galaxies could be weighted by their colour etc. Given that, as well as the colour-density relation, there is a morphology-density relation, it is natural to ask which came first i.e. is colour fundamentally tied to density, and the morphology-density relation just arises because colour and morphology are correlated, or vice versa, or both/neither? They found that, at fixed colour, a correlation function marked by morphology had no enhancement. However, at fixed morphology (and luminosity), a correlation function marked by colour had substantial enhancement - suggesting that the colour-density relation is the more fundamental relation. This agrees with my results that correlation function measurements can give meaningful insight into the colour-density relation, as well as being reassuring that colour-density is the more fundamental relation to investigate (seeing as ground based surveys in general have comparatively little morphological information at high redshift). In addition, although at substantially different redshifts ($z \sim 0$) to this study, they have some results (fig 9 in Skibba et al., 2009) qualitatively similar to my $z \sim 1.2$ measurements of enhancement *only* existing on small (< 0.5Mpc) scales. They found that their red galaxies have the same clustering power regardless of morphology on larger scales, but that on smaller scales clustering is enhanced for ellipticals. Further work is needed to see if this morphological effect is linked to my results, or if it is merely a chance result of the two works relative choices of colour/sSFR division etc.

My results in terms of environmental dependence appearing in the 1-halo term agree well with Hearin et al. (2016a), who suggest that there should be little to no conformity at high redshift, 1-halo conformity at intermediate redshifts, and be present in both 1-halo and 2-halo scales at redshifts towards $z \sim 0$. This is very similar to the picture I see of no enhancement at high redshift, enhancement on small scales at intermediate redshifts, and enhancement on all scales at the lowest redshifts.

Darvish et al. (2016) is perhaps the most similar study to this chapter. They look at

galaxies with photometric redshifts up to $z \sim 3$ in 2 sq deg of the COSMOS field, with a limiting K-band magnitude of 24. They use Voroni tessellations to give their density measure (in the manner of Darvish et al., 2015), and find that, at fixed stellar mass, sSFRs are up to two orders of magnitude lower in denser regions, but that at higher redshifts ($z > 1$) this dependence essentially vanishes. This is in qualitative agreement with my result that the sSFR-density relation predominantly comes into action at lower redshifts.

4.5.3 Future Development of the Approach

As discussed in section 4.5.1, the modelling discussed here relies on some approximations that start to fail in some regimes probed in this paper. In addition there are still degeneracies in the modelling, and Scenario 3 is less revealing than Scenario 1 or 2 as it does not allow discrimination between quenching from location in the halo, and quenching from proximity to galaxy. Both of these mean that a logical next step is to incorporate a quenching model for all halo masses and stellar masses, so that both central-satellites and satellite-satellite pairs can be incorporated, from all halo masses. As more data become available, and the model is developed, I believe that a more robust understanding of interactions between galaxies within a halo will be obtainable, and more definitive statements about the distribution of star formation in galaxies within a halo will become possible.

More complex scenarios, where the quenching mechanism is more complicated than considered here can be probed by simulating galaxy locations in halos, as I do for the auto-correlations of Scenario 1. The massive galaxy and less massive galaxy have locations in the halo drawn from a distribution that traces the halo profile. Then according to equation 4.12, the lower mass galaxy is determined probabilistically to be passive or star-forming. This allows us to construct essentially D_1D_2 of the correlation function, and therefore a 1-halo term corresponding to a physical quenching process. This can be used to measure the effects of multiple quenching sources in the same halo etc.

In general, for a model that specified galaxy occupation numbers and interaction terms, all possible auto and cross correlation 1-halo terms can be constructed in this Monte Carlo manner: 1) sample halo mass from the HMF 2) sample how many galaxies are in the halo (given an average occupation number as a function of halo mass and assuming Poisson) 3)

sample locations for the galaxies within the halo 4) determine star forming/AGN/galaxy properties of choice for galaxies based on probabilistic model of galaxy interactions 5) record the number of galaxies of each species and all pairs of spatial separation. This allows one to construct all auto and cross 1-halo terms for arbitrarily complicated models of the halo environment. The parameters that describe the halo environment interactions can then be explored with MCMC techniques. This approach is much more computationally challenging than the cases for which analytic expressions are available, but may provide a useful intermediary between the cases described by analytic expressions, and a full-blown cosmological simulation. Another possible approach would be to 'graft' a galaxy occupation and interaction model on a dark matter only simulation as per the comprehensive HALOTOOLS code described in Hearin et al. (2016b).

Once the physical model is used to construct the 1-halo term, it can be added to a 2-halo term, simply constructed by scaling the dark matter correlation function by the galaxy biases of the two samples. This spatial cross-correlation function can then be compared with observations by projecting to an angular correlation function etc.

I note that understanding the cross-correlation function is important for other applications e.g. Croft (2013) consider a very similar scenario to us of high-mass galaxies cross-correlated with low-mass galaxies, for the purpose of detecting gravitational redshift-space distortions (see section 2.1.1.1).

4.6 Conclusions

I have used data from the VIDEO survey to investigate the processes which results in environmental quenching and galactic conformity. In particular I have studied the cross-clustering of galaxies with the two point cross-correlation function up to $z \sim 2$, varying the stellar mass and star formation rates of the galaxies. Building on the JHOD formalism introduced by Simon et al. (2009) I used MCMC methods to explore the probability of a galaxy being quenched at different redshifts, different radial locations within a halo, and when the central galaxy was star forming or passive.

Key conclusions:

- Cross correlations between galaxy samples of different star formation rates contain information about the role of environment in determining whether a galaxy is star forming or passive
- Measurements of the cross-correlations suggest that the sSFR-density relation is non-existent/weak at $z \sim 2$
- The sSFR-density/environmental quenching can be seen to emerge at $z \sim 1$ and grows increasingly important towards $z = 0$ as galaxies are increasingly more likely to be quenched in group environments over the field
- Lower mass galaxies appear to be more likely to be quenched in the vicinity of a massive passive galaxy, than a massive star forming galaxy of the same stellar mass
- There appears to be little radial dependence for being quenched within a halo at lower redshifts, supportive of the quenching of accreted galaxies being caused by the cessation of fresh inflows of gas, as opposed to tidal stripping
- A quenching mechanism within a halo can be added into the HOD framework, and may be a promising approach for future environment studies, as well as being important to understand for advancing the HOD programme of understanding the galaxy-halo connection

Planned future work will develop the formalism so that all halo masses are incorporated and properly controlled for. I suggest that in the future, this more sophisticated model, combined with Bayesian model selection methods may allow one to discriminate between different forms of environmental quenching e.g. whether environmental quenching is really associated with halo mass, location in the halo, proximity to another galaxy or the presence of an active galactic nucleus.

Chapter 5

The environment and host haloes of the brightest $z \sim 6$ LBGs

This Chapter closely follows work by Hatfield, Bowler, Jarvis and Hale (2017, submitted).

5.1 Introduction

5.1.1 Lyman-break Galaxies

The study of Lyman-break Galaxies (LBGs) is a long established probe of the high-redshift Universe (the first few billion years), with samples of many hundreds of star-forming galaxies now known to $z \sim 10$ (e.g. Oesch et al., 2016; Bouwens et al., 2016; McLeod et al., 2016). LBGs are particularly useful as it is possible to establish their photometric redshift to reasonable accuracy in a luminosity regime where spectroscopic confirmation is challenging (e.g. Pentericci et al., 2014). The neutral gas in the inter-galactic medium (IGM) is essentially opaque to photons with wavelengths shorter than the 'Lyman Break' (1216Å, in the far ultraviolet). The source therefore appears faint bluewards of this wavelength, but retains its original luminosity redwards, creating a sharp drop in luminosity. When this spectrum is then redshifted, the location of the break provides a clear spectral feature with which to select galaxies to high-redshifts using broad-band filters.

The technique, originally developed in the early 1990's (Guhathakurta et al., 1990;

Steidel & Hamilton, 1992; Steidel et al., 1996) in the context of $z \sim 3$ galaxies, where the Lyman break is shifted into visible wavebands, first started providing large numbers of sources with the *Hubble Space Telescope* (HST) in the late 1990's and 2000's (e.g. Giavalisco, 2002; Bouwens et al., 2007; Dunlop et al., 2013). More recently, the approach is being used to push scientific boundaries at $z \sim 6-9$ where the break is shifted into the near-infrared (see Stark, 2016 for a review). Wide field surveys like the United Kingdom Infrared Telescope (UKIRT) Infrared Deep Sky Survey (UKIDSS), in particular the Ultra Deep Survey (UDS, Hartley et al., 2013), and more recently public surveys on the Visible and Infrared Survey Telescope for Astronomy (VISTA) such as the UltraVISTA survey in the COSMOS field (McCracken et al., 2012) and the VISTA Deep Extragalactic Observations (VIDEO) survey (Jarvis et al., 2013) give access to the deep NIR images of the sky needed for detecting statistically significant samples of the brightest LBGs, which has led to advances in the understanding of their star formation rates and number densities beyond the break in the luminosity function, where the most luminous galaxies reside.

A key observable that can be calculated for LBG surveys (and galaxy surveys in general), is the luminosity or mass function, the comoving number density of galaxies as a function of absolute luminosity or stellar mass (see Chapters 1 and 2, and Johnston, 2011 for a review). Measuring and understanding the evolution of luminosity functions with redshift allows us to trace the build-up and evolution of galaxies through cosmic time (Madau & Dickinson, 2014); is a key way to compare cosmological simulations of structure formation to observations (Lacey et al., 2016; Clay et al., 2015); and can be readily linked theoretically to the dark matter equivalent, the halo mass function (e.g. Murray et al., 2013 for a review of current constraints). Luminosity functions are typically observed to have the form of a Schechter function: $n(L) = \phi^*(L/L^*)^\alpha \exp(-L/L^*)$ (Schechter, 1976). In this parametrisation, α describes the power law behaviour of number density at the low-luminosity end, L^* is the transition luminosity to the high luminosity exponential cutoff, and ϕ^* is a normalisation. The rest frame UV luminosity function for $z \sim 4-8$ has been determined by several studies (e.g. recent work by McLure et al., 2013; Bouwens et al., 2015; Finkelstein et al., 2015; Bowler et al., 2015) with broad agreement. The highest redshift constraints on the LBG luminosity function are currently at $z \sim 9-10$ (Bouwens et al., 2015, 2016; McLeod

et al., 2016).

5.1.2 Clustering

Galaxies are formed and live in dark matter halos, and the environment of the host halo is believed to be of critical importance for the formation of the resident galaxies (Cooray & Sheth, 2002). One way of obtaining information about the galaxy-halo connection is 'abundance matching' - matching the galaxy comoving number density value to the halo mass that is predicted to have the same number density by theoretical considerations of the halo mass function/N-body simulations (see Chapter 1). Abundance matching however can only ever give an incomplete account of the connection due to three complications. Firstly, halos can host multiple galaxies, this can be partially mitigated through *sub*-halo matching (Moster et al., 2010), but this assumes that the occupation statistics are the same for sub-halos and isolated halos. Secondly, scatter in the halo mass to galaxy mass/luminosity relation is not captured by abundance matching. Finally, variations in observational properties of a single population can bias results, in particular orientation or temporal effects e.g. if a given population has a different appearance 10 percent of the time, a straightforward abundance matching will erroneously place these sources with a different appearance in more massive halos as they are rarer, even though they are the same object as the underlying population. This is of particular interest in this analysis as some authors suggest that LBGs can have highly variable star formation rates over time (e.g. Stark et al., 2009). For this reason, other probes of the galaxy halo connection are needed. Information from lensing is very effective, either strong (e.g. Jullo et al., 2007) or weak (galaxy-galaxy lensing, e.g. Coupon et al., 2015; Mandelbaum et al., 2014), but requires the background sources to be at even higher redshift than the lenses, making it unfeasible for high ($z > 2$) galaxies.

I therefore use the approach of measuring the clustering of the galaxies alongside the number counts (see Chapter 2). This can then be linked to models/our theoretical understanding of structure formation to estimate the typical environment of the galaxies - see Chapters 2 and 3. In the high-redshift regime, the HOD model has recently been applied to low-luminosity LBG galaxies at $z = 4 - 7$ by Harikane et al. (2016). It is crucial however to understand the relationship at the massive/most luminous end, as this is where AGN-

driven feedback may have an important role (Silk, 2010). There are preliminary hints that redshifts of $z = 6 - 7$ may mark the onset of quenching (Bowler et al., 2014, 2015), so this is a vital time period for galaxy evolution in the history of the Universe.

5.1.3 Probes of Reionization

As well as being a crucial period for galaxy formation (see Shapley, 2011 for a review), understanding large-scale structure/clustering at $z = 5-8$ is also important cosmologically, in particular for our understanding of *reionization*. In the standard cosmological model, the Universe was initially an ionised plasma, that during *recombination* at $z \sim 1000$ cooled sufficiently for protons and electrons to combine to form neutral atoms. However this medium must have been reionized by the first stars and galaxies at some point between then and $z \sim 6$ to produce the ionized intergalactic medium seen today (see Zaroubi, 2013; Natarajan & Yoshida, 2014 for an observational and theoretical review respectively). However there is still debate in the literature about which sources had the energy to do this e.g. massive galaxies, AGN, or something else (Grissom et al., 2014; Robertson et al., 2015; Bouwens et al., 2015). Furthermore, most models of reionization are 'patchy' e.g. non-instantaneous, with some parts of the Universe being reionised earlier than others (Becker et al., 2015; Doré et al., 2007). The current best constraint on the average reionization redshift from the Planck mission (based on the measured optical depth) is $z = 7.8 - 8.8$ (Adam et al., 2016), with many probes of the epoch (e.g. Pentericci et al., 2014; Becker et al., 2015) suggesting that some parts of the Universe could still be undergoing reionization by $z \sim 6$. Uncovering the cause and nature of reionization is likely to require understanding how the reionization power spectrum interweaves with galaxy large scale structure, so it is essential to build up our understanding of the large scale structure (LSS) of the galaxy population at these redshifts. For example, McQuinn et al. (2007) suggest that differences in the clustering of LBGs and Lyman-α Emitting galaxies (LAEs) could give an insight into the possible 'patchy' nature of reionization. The Lyman-α line is suppressed if the source is in a largely neutral region which biases the observations of LAEs towards large ionized HII regions. The result is a larger 'observed' clustering for LAEs than the 'intrinsic' clustering of the underlying objects - effectively neutral regions obstruct the line of sight in

a way that enhances the clustering of LAEs. LBGs and other probes of the high redshift galaxy population however do not receive such an effect on their clustering, so a boost in the clustering of LAEs relative to LBGS (properly controlling for other variables) could be indicative of the form of reionization. Such an effect is yet to be conclusively measured, e.g. Ouchi et al. (2010) find little evidence that reionization could be detected at $z = 6.6$ with 207 LAEs observed with the Subaru telescope, but this approach and others like it are likely to give improved constraints to the nature of the epoch of reionization over the coming years.

5.1.4 Objectives

LBG studies can be informally divided into analyses of 'faint' galaxies (in extremely deep, but narrow surveys), and 'bright' galaxies (in slightly less deep, but wide surveys). Harikane et al. (2016) provide an analysis of the clustering of relatively faint LBGs found within HST deep surveys at $z = 4 - 7$. In this Chapter I extend these measurements to brighter luminosities by utilising wider area surveys. To do this, I measure and model the clustering of the Bowler et al. (2015) high luminosity $z \sim 6$ LBG sample, which covers the degree-scale UDS (Furusawa et al., 2008; Lawrence et al., 2007) and UltraVISTA (McCracken et al., 2012; Laigle et al., 2016) fields. A clustering analysis of a subset of the UDS sample has been performed in McLure et al. (2009), who modelled the correlation function with a single power law, concluding the sources are in dark matter haloes of masses $10^{11.5-12} M_{\odot}$. In this study I perform a similar analysis, but extend to a full HOD model, including an additional field and using deeper data. Using this larger sample, I then discuss my results in the context of feedback processes, models of structure formation, and cosmic variance at high redshift. While samples of bright galaxies do exist at $z > 6.5$ (Bowler et al., 2015), they are too small to provide constraints on the clustering, and hence I limit my analysis to $z \sim 6$.

The structure of this Chapter is as follows. In Section 5.2 I describe the sample of LBGs used in this study. In Section 5.3 I discuss how I measured the correlation function in the sample and constructed the halo occupation distribution models and fitting process. The results are presented in Section 5.4. In Section 5.5 I discuss the results, linking them to

the literature, and interpreting the cosmic variance between the fields in light of the new measurements.

5.2 Data and Sample Selection

In this study I use the high luminosity LBG sample of Bowler et al. (2015). Deep optical and infrared data (spanning wavelengths of $0.3 - 2.5\mu$m) across two fields (see Fig. 5.1) was used to select the sample; I summarise the observations and selection criteria below, but see Bowler et al. (2015) for a more in depth description.

5.2.1 UltraVISTA/COSMOS

UltraVISTA (McCracken et al., 2012; Laigle et al., 2016) is the deepest of the 6 public surveys on the VISTA telescope, providing $YJHK_s$ near infra-red data covering the Cosmic Evolution Survey (COSMOS) field (Scoville et al., 2007). The 'paw-print' focal plane of VISTA and the survey observing strategy give a continuous 'deep' field, with discontinuous 'ultra-deep' stripes across it that receive more observing time. Bowler et al. (2015) also used u^*, g, r and i optical data from the T0007 release of CFHTLS in the D2 field, as well as Subaru/SuprimeCam z'-band imaging. The maximal area of overlap of these datasets is in the one square degree of CFHTLS, of which 0.62 deg^2 has ultra-deep UltraVISTA data, and 0.38 deg^2 shallower UltraVISTA. The majority of the sample is in the ultra-deep field and hence for the purpose of this work I only use the ultra-deep 0.62 deg^2 (see Fig. 5.1).

5.2.2 UDS

For the UKIDSS UDS field, Bowler et al. (2015) used B, V, R, i and z' data from the Subaru XMM-Newton Deep Survey (SXDS, Furusawa et al., 2008), and J, H and K band data from the DR10 of the UKIDSS UDS (Lawrence et al., 2007). Again separate z'-band data from Subaru/SuprimeCam was obtained, and in addition, Y band data from the VIDEO survey (Jarvis et al., 2013) was also used. The total overlapping area is 0.74 deg^2 (see Fig. 5.1).

5.2.3 Candidate Selection

Again, Bowler et al. (2015) describes the full sample selection, but I summarise the process here. Sources were detected with SExtractor (Bertin & Arnouts, 1996), and photometric redshifts were determined with LePhare (Arnouts et al., 1999; Ilbert et al., 2006). Contaminant populations (low redshift interlopers and brown dwarfs) were removed in the SED fitting process. This leaves 156 and 107 $5.5 < z < 6.6$ galaxies in the UltraVISTA and UDS fields respectively. The UltraVISTA field was found to have a higher surface density than the UDS field (by a factor of ~ 1.8); potential causes for this, including lensing and cosmic variance are discussed in section 7 of Bowler et al. (2015).

This process gives in total 263 LBGs in the range $-22.7 < M_{UV} < -20.5$ with $5.5 < z < 6.5$ over 1.35 deg^2. I take my final sample as all 161 sources with $M_{UV} < -21.125$, as the sample completeness drops off rapidly faintwards of this value, as discussed in Bowler et al. (2015), see their figure 6, but is fairly constant with magnitude brightwards of this value.

5.3 Correlation Functions and HOD Modelling

In a similar vein to Chapter 3, I calculate the angular correlation function as discussed in Chapter 2. I use the Landy & Szalay (1993) estimator, estimate the correlation function continuously as per Hatfield et al. (2016), and include the integral constraint in the model as per Beutler et al. (2011). I do not use the photometric redshift probability distribution function weighting system described in Chapter 2 here, as all the sources are within the redshift range under consideration to a high degree of confidence.

For the construction of the random catalogue I created a mask over the fields to exclude image artefacts and foreground stars, see Figure 5.1. Five galaxies in the UDS field were found to be within the masked area. Although the mask may be a little conservative, it is likely that the measurements of clustering in the vicinity of these sources will be heavily biased by the presence of the artefact being masked, so I do not use these five galaxies when calculating the correlation function (although it makes very little difference to my analysis).

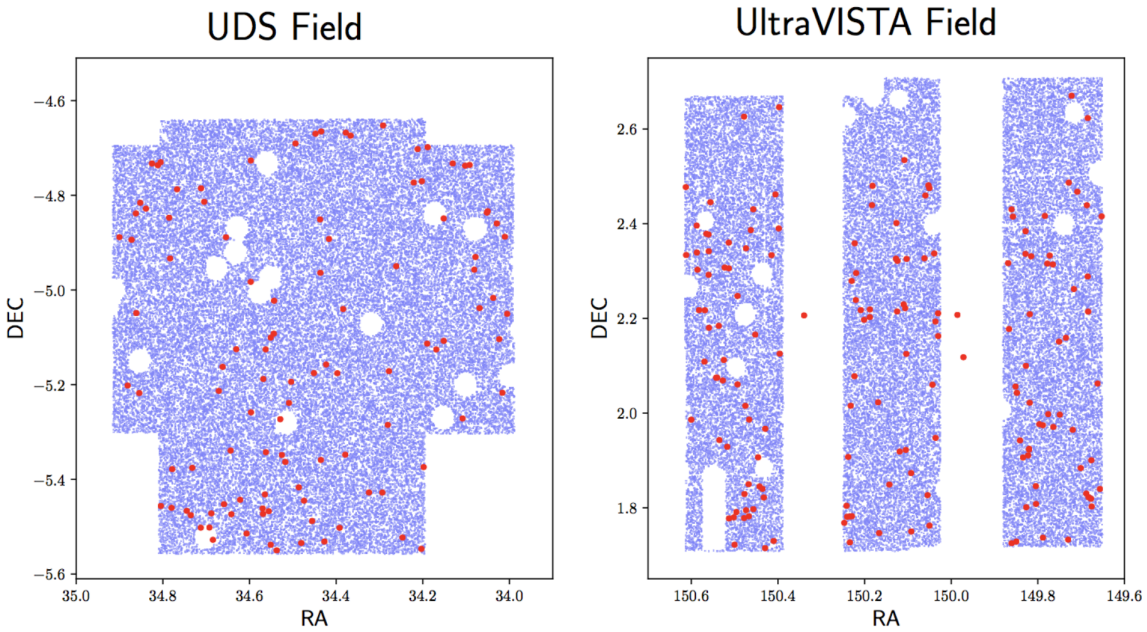


Figure 5.1: The geometry of the UDS and UltraVISTA fields. The red points are the galaxy locations from Bowler et al. (2015). The blue points are the random points chosen to cover the fields used for the construction of the random galaxy catalogue for the calculation of the correlation function. The three galaxies in the UltraVISTA field that are not surrounded by blue points are the $z = 6$ sources detected in the deep (as opposed to ultra-deep) part of the UltraVISTA field, that I do not include in this study. The overall shape of the fields is predominantly determined by the part of the sky that the multiple different surveys overlap in. The small scale gaps and holes are foreground stars and detector artefacts etc. See figures 1 and 2 of Bowler et al. (2014) to see how the irregular footprints arise from the intersection of the sky patches covered by different surveys.

5.3.1 Halo Occupation Distribution modelling

Similarly to Chapter 3, I use the 5 parameter HOD model of Zheng et al. (2005) that I describe in Chapter 2, assuming a Navaro-Frenk-White profile (Navarro, Frenk, & White, 1997) and a Tinker et al. (2010) bias model. I use the HALOMOD python package to calculate model correlation functions and derived parameters. In this Chapter however I use the halo mass function of Behroozi et al. (2013), which is a high redshift calibration of the Tinker et al. (2010) HMF.

Although this is the largest sample of bright LBGs at these redshifts, it is still only a comparatively small sample for HOD modelling. Thus, in order to reduce the number of parameters in the model (six once duty cycle is included, see Section 5.3.2), I fix some as

functions of others.

As per Harikane et al. (2016) I fix $\sigma_{\log M} = 0.2$. The assumptions that go into this choice however are based on results at much lower redshifts (Kravtsov et al., 2004; Zheng et al., 2005; Conroy et al., 2006), which do not necessarily hold at these early times, when the luminosity-halo scatter is fairly unconstrained. Indeed in Chapter 3 I found a scatter of ~ 0.6 at $z \sim 1$. Fortunately for the purposes of this chapter (unfortunately from the perspective of using clustering to infer the scatter) the 2-point statistics have very little dependence on the scatter. Hence the conclusions do not alter dramatically with choice of $\sigma_{\log M}$, and so I fix it as the same as the Harikane et al. (2016) value for ease of comparison.

Additionally I fix $\alpha = 1$; this is both the fiducial value (it is logical to expect that once in the most massive halo regime that the number of satellites scales linearly with the halo mass, as the bulk of the halo mass will have been accreted), as well as the result found by most measurements at moderate ($z < 2$) redshift (e.g. Hatfield et al., 2016; Coupon et al., 2012).

I investigate the consequences of allowing various parameters to be fixed or free in the fitting process. Again as per Harikane et al. (2016), if not free, M_1 and M_0 are fixed as functions of M_{min} following the $z = 0 - 5$ results of Conroy et al. (2006):

$$\log M_1 = 1.18 \log M_{min} - 1.28, \tag{5.1}$$

$$\log M_0 = 0.76 \log M_1 + 2.3. \tag{5.2}$$

5.3.2 Duty Cycle

The role of a duty cycle (DC) is the main difference to be incorporated when modelling rest-frame UV-selected LBG galaxies at high redshift compared with studies in the local Universe based on stellar mass. Clustering analyses of LBGs typically find that there is a mismatch between the measured number density, and the number density implied by the clustering (Ouchi et al., 2010). This is in agreement with the current understanding of galaxies at these redshifts that suggest that star formation may be highly episodic (e.g.

Stark et al., 2009). Typically the occupation statistics model obtained by fitting *only* to the clustering will give a larger comoving number density than is observed in the luminosity function. This discrepancy is typically explained by invoking a *duty cycle*, that the observed LBGs have luminosities that vary dramatically in time, and are being observed only when in a bright phase. This illustrates the importance of understanding clustering alongside the number counts. With a duty cycle of 10 percent (i.e. it is only in its bright phase 10 percent of the time), the underlying galaxy appears 10 times rarer than it actually is. A straight abundance matching in this scenario would then mistakenly put them in rarer, and thus more massive, halos.

Without incorporating the duty cycle, the implied comoving number density is the mean number of galaxies in a halo, multiplied by the halo mass function, integrated over all halo masses. This number is then multiplied by the duty cycle to give the model comoving density:

$$n_{gal} = \mathrm{DC} \times \int_0^\infty \mathrm{HMF}(M_h) \times \langle N_{tot}(M_h) \rangle \mathrm{dM}_h, \tag{5.3}$$

where HMF is the halo mass function and DC is the duty cycle.

5.3.3 MCMC Fitting

I use EMCEE[1] (Foreman-Mackey et al., 2013) to provide a Markov Chain Monte Carlo sampling of the parameter space to fit the correlation function. I use a likelihood of:

$$\chi^2 = \frac{[\log n_{\mathrm{gal}}^{\mathrm{obs}} - \log n_{\mathrm{gal}}^{\mathrm{model}}]^2}{\sigma_{\log n}^2} \tag{5.4}$$

$$+ \sum_{i,j} [\omega^{\mathrm{obs}}(\theta_i) - \omega^{\mathrm{model}}(\theta_i)][C_{i,j}^{-1}][\omega^{\mathrm{obs}}(\theta_j) - \omega^{\mathrm{model}}(\theta_j)], \tag{5.5}$$

where $n_{\mathrm{gal}}^{\mathrm{obs}}$ is the observed galaxy number density, $n_{\mathrm{gal}}^{\mathrm{model}}$ is the model galaxy number density, $\sigma_{\log n}$ is the error on the log of the number density including both Poisson noise and cosmic variance, θ_i are the angular scales I fit over, ω^{obs} is the observed angular correlation function, ω^{model} is the angular correlation function of a given model, and $C_{i,j}$ is the

[1] http://dan.iel.fm/emcee/current/

covariance matrix of the measurements of the correlation function from the bootstrapping.

When the parameters are free, I use a uniform prior over $10 < \log_{10}(M_{\rm min}/M_\odot) < 13$, $\log_{10}(M_{\rm min}/M_\odot) < \log_{10}(M_1/M_\odot) < 14$ (uniform in log space) and $0 < {\rm DC} < 1$. I use 20 walkers with 1000 steps, which have starting positions drawn uniformly from the prior.

I use 500,000 random data points in this study. As per Chapter 3 and Chapter 4, I use bootstrap resamplings to estimate the uncertainty at the 16th and 84th percentiles of the resampling. For $n_{\rm gal}^{\rm obs}$, I use the value obtained when integrating the luminosity function brightwards to infinity from $M_{UV} = -21.125$ (as opposed to the number obtained by dividing the number of sources by the volume probed) as the luminosity function already has incompleteness factored in. This equates to $n_{\rm gal}^{\rm obs} = 4.1\times10^{-5}{\rm Mpc}^{-3}$, for the fields combined. This is from the best-fitting double power law model in Bowler et al. (2015). Using the best-fitting Schechter function gives the marginally lower value of $n_{\rm gal}^{\rm obs} = 3.8\times10^{-5}{\rm Mpc}^{-3}$. Changing from one value to the other does not impact my conclusions. The main sources of incompleteness are blending with foreground sources, and misclassification of true $z \sim 6$ LBGs as dwarf stars or lower redshift contaminants (see Bowler et al., 2015).

5.4 Results

5.4.1 Clustering Measurements

Fig. 5.2 shows the angular correlation function of the full sample over the range $10^{-3} < \theta/{\rm deg} < 10^{-0.5}$, estimated with both the binning approach and the kernel smoothing method. They (as expected) agree well, and produce the familiar approximate power law $\sim \theta^{-0.8}$, although the kernel smoothing method is able to cope better with bins that contain a small number of pairs. For the rest of this analysis, I take the value of the smoothed correlation function, at the ten angular scales calculated for the bins, as my final measurements[2].

In general, measurements of clustering at different scales will be covariant as individual

[2]With the continuous estimation of the correlation function, one can in principle extract an estimate of the correlation function at an arbitrary number of angular scales in the range probed. However this gives dramatically diminishing returns as adjacent measurements would be increasingly covariant e.g. one could take the estimate of the correlation function at 1000 points in the angular range for which I estimate the correlation function, but adjacent points would be almost perfectly correlated and no extra information would be gained.

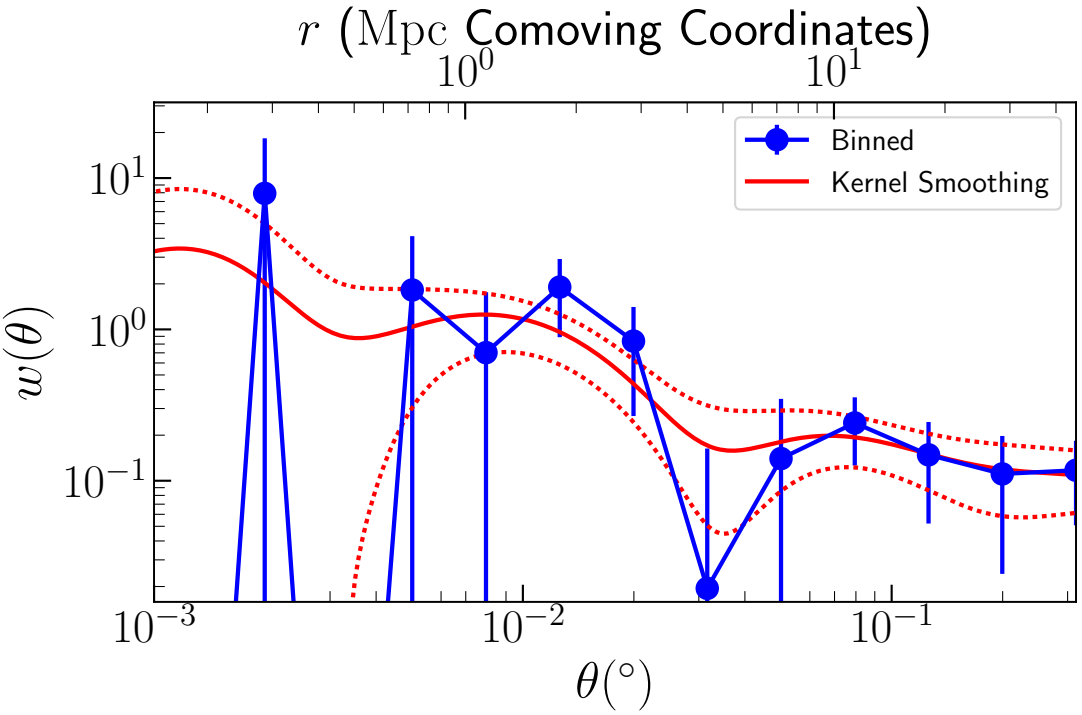


Figure 5.2: The angular correlation function for the sample of bright ($M_{UV} < -21.125$) $z \sim 6$ LBGs from Bowler et al., (2015). The figure shows the correlation function estimated both with a binning method (blue points), and a kernel smoothing method (red curve, with dotted lines showing uncertainty). Where the binned correlation function dips to negative values corresponds to where there were no galaxy pairs in the bin.

galaxies contribute multiple times to DD, usually at different scales. Furthermore, extra care with covariances is needed when using the kernel method, as a given galaxy pair contributes at a range of scales (this can be mitigated by picking measurements larger than the smoothing scale, but is important to keep track of here as I am in a low-data regime). Therefore I also construct the covariance matrix from the bootstrapped samples for my measurements, in order to account for these covariances in the fitting process. I show the correlation matrix (the covariance matrix with each value normalised by the standard deviation of each measurement) in Fig. 5.3. Not taking covariances into account would be the equivalent of ignoring the off-diagonal values, which are non-negligible for this study, in particular at the large and very small scales.

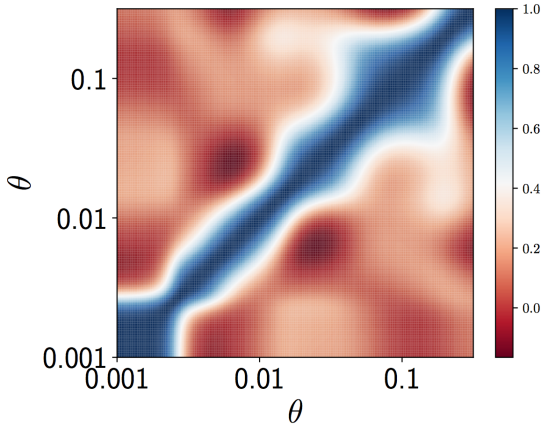


Figure 5.3: The correlation coefficients (covariance normalised) of my measurements. Blue values are positive correlations, red values are negative values. The blue diagonal corresponds to the standard deviation measurements (a random variable is always perfectly correlated with itself).

5.4.2 Modelling Results

I carry out several MCMC fits to the data with the HOD model as per Section 5.3.3, with four variations. The four scenarios considered were:

- M_{min}, M_1 and DC free (e.g. 1-halo and 2-halo amplitudes and number counts all free)
- M_1 fixed as function of M_{min} as per Section 5.3.1 (e.g. 1-halo amplitude as fixed function of 2-halo amplitude)
- M_1 fixed as function of M_{min}, DC = 0.6 (e.g. 1-halo amplitude as fixed function of 2-halo amplitude, duty cycle fixed at the Harikane et al., 2016 value)
- M_1 fixed as function of M_{min}, DC = 1 (e.g. 1-halo amplitude as fixed function of 2-halo amplitude, no duty cycle - galaxies 'on' at all times)

In addition to these four models I also compute:

- Halo masses for the most straightforward abundance matching scheme e.g. M_{min} for a sample of galaxies above a given luminosity threshold is the halo mass such that the comoving number density of halos greater than that mass is equal to the comoving number density of the galaxy sample (see Table 5.2)

- Galaxy bias from a pure bias model e.g. fit for b where $\xi_{\rm model} = b^2 \xi_{\rm DM}$.

The results from these 6 models are shown in Tables 5.1 and 5.2. Fig. 5.4 shows the data, and the best-fit models. I show the posterior from the M_{min}, M_1 and DC free fit in Fig. 5.5 for illustrative purposes. It is clear that the amplitude of the correlation function is roughly two orders of magnitude larger than the dark matter correlation function in the linear regime, corresponding to a very high bias. Most of the models suggest that $M_{min} \sim 10^{11.5} M_{\odot}$ e.g. these galaxies are hosted by halos of that mass and above. It is also clear that the satellite fraction is at most a few percent, which suggests that at most 5-6 galaxies in the sample are satellites (in the scenario that these sources were the same underlying population as the lower luminosity LBGs, the satellite fraction could have been higher as a non-trivial portion would have been from halos hosting multiple galaxies).

It becomes more difficult however to make statements beyond these basic claims because the HOD fits are only of moderate quality (see the χ^2 values in Table 5.1)[3]. In general, fits decreased in quality as more parameters were fixed, as expected. I discuss the tensions more in section 5.5.1, but summarise the results of each model here. The M_{min}, M_1 and DC free model is free to go to high masses until tension between model and measured number density stop it from going higher. This model can also take M_1 extremely high, to bring the amplitude of the small scale clustering down to match the data. Models that do not have M_1 free cannot vary their small scale behaviour freely. This forces their halo masses down, as the small scale behaviour grows rapidly with M_{min}; if they went higher the disagreement on small scales would become much larger. When DC is free (and M_1 is fixed as a function of M_{min}), the model actually prefers to go even lower than the abundance matching halo mass, and uses the duty cycle to reach agreement with the number counts. However when these models have the duty cycle fixed, they cannot do this, so the trade off between agreeing with small scale clustering and the number counts sets the halo mass. For DC = 0.6 to agree with the observed number counts, the intrinsic number counts must be higher than for DC = 1, forcing the model to prefer slightly lower halo masses. Conversely

[3]N.B. The moderate quality of the fits is responsible for the conspicuously small error bars on some of my results in table 5.1 and figure 5.7 - the tension between different parts of the data forces the best fit to be at a point between the two conflicting measurements to a high level of precision.

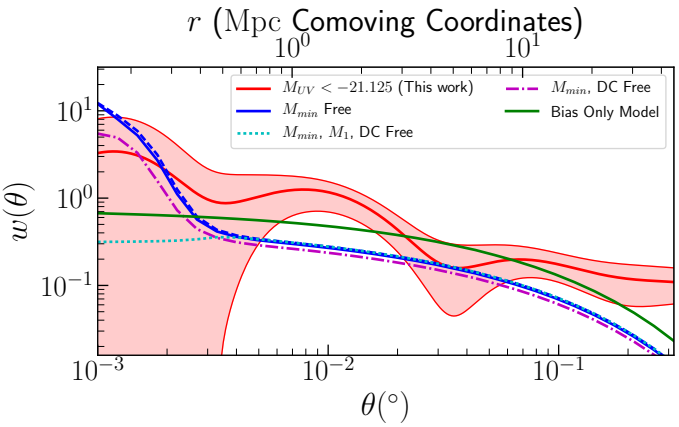


Figure 5.4: Comparison of the measurements of the correlation function (red curve and shaded area) with the five different models I fit. The two blue curves correspond to models with only M_{min} free, the dashed curve has DC = 1 and the full line DC = 0.6. On linear scales all models are very similar, apart from the bias only model, as the number density constraint is restricting the model from going too high. Only the M_1 free model allows the small-scale amplitude to vary independently.

the 'pure bias' model was able to fit the clustering data well. This suggests that the reason for the poor quality fit is a mismatch between the clustering and the number counts - halos of the halo mass implied by the bias are far rarer than the observed galaxies are[4], which is problematic - although part of the raison d'etre of HOD schemes is to understand how multiple galaxies can occupy the same halo, which would allow the number of galaxies to be greater than the number of halos, this is, as discussed, a few percent effect, as opposed to a factor of ten effect. This is the opposite problem to what the duty cycle is invoked to solve - duty cycles in clustering studies of LBGs solved the issue of number counts being lower than implied by clustering, the problem here is the number counts are higher. In addition, the fact that no 1-halo term emerges is slightly anomalous. I note that Harikane et al. (2016) use fitting formulae of the HMF in Tinker et al. (2010) directly without the normalization constraint, which overestimates the abundance by a factor of 1.7 at $z = 4$ (Y. Harikane 2017, private communication). So the results of Harikane et al. (2016) are likely more consistent with a duty cycle of 1 (rather than the fixed value of 0.6 that they use in their analysis).

[4]This problem would have been even worse if the HMF had been taken from Tinker et al. (2010) without the high redshift correction of Behroozi et al. (2013).

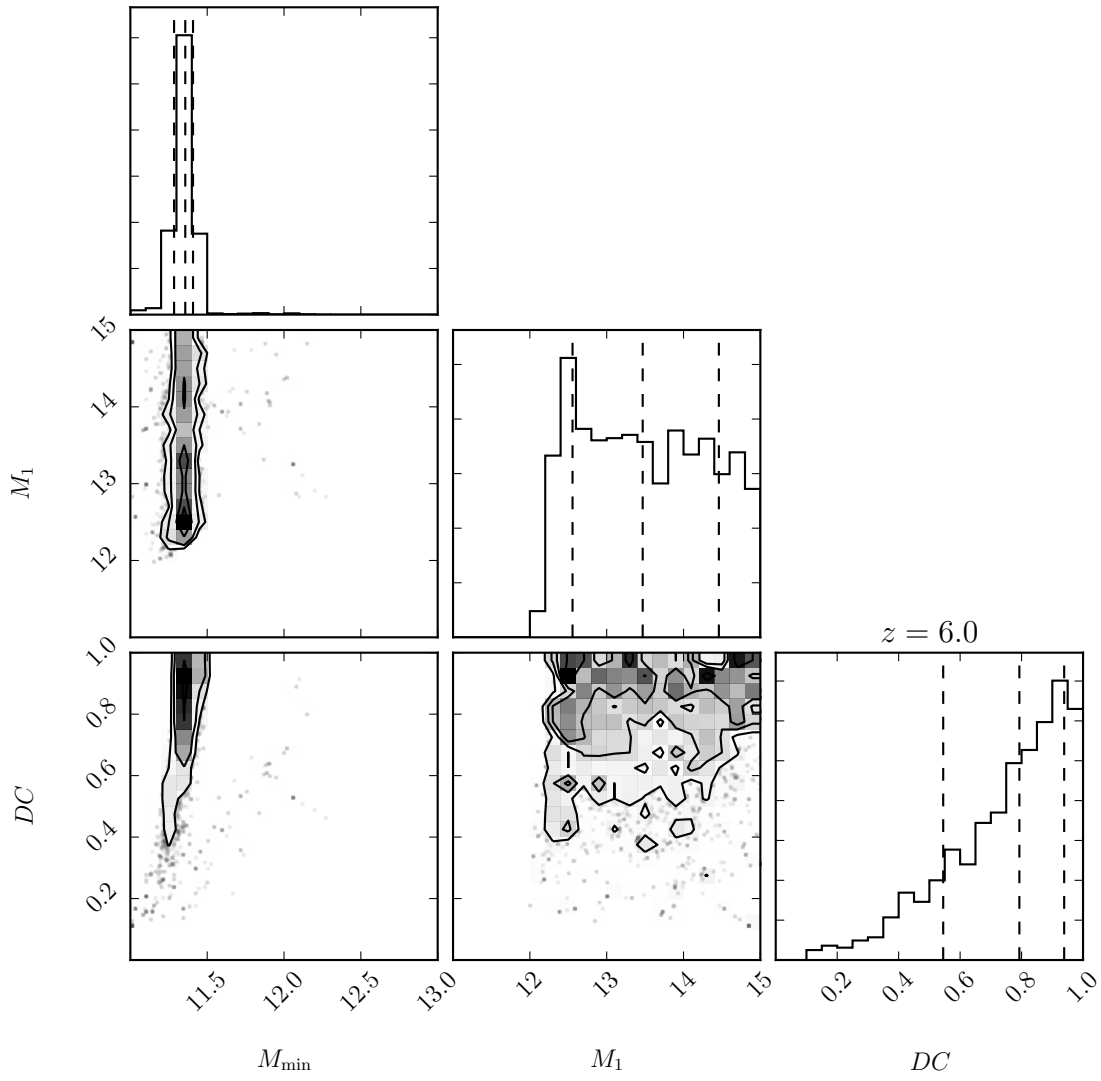


Figure 5.5: Triangle plot of the posterior from the MCMC fitting for the HOD model with M_{min}, M_1 and DC free (masses in log base ten Solar mass units). Dashed lines on the one dimensional single parameter plots are 16th, 50th and 84th percentiles.

Table 5.1: Fitted HOD parameters from the MCMC. Also shown are the corresponding satellite fractions ($f_{\rm sat}$) and galaxy biases (b) and fit reduced χ^2 of the samples. Quantities in brackets are either fixed in the model, or fixed as a function of other parameters in the model. Masses are in Solar mass units (log base ten). Note that values and error bars quoted are the 16th, 50th and 84th percentiles of the posterior, as opposed to the peak values. This makes very little difference apart from the posterior for the duty cycle value for the M_{min}, M_1 and DC free model, which is peaked at DC = 1 and hence only has one tail, see Fig. 5.5. The lower luminosity parameter values are taken directly from Harikane et al. (2016), apart from satellite fraction, which I calculate.

Model	M_{UV}	$\log M_{min}$	$\log M_1$	$\log M_0$	α	σ	DC	$10^2 f_{sat}$	b	χ^2/d.o.f.
M_1, DC free	-21.125	$11.53^{+0.05}_{-0.07}$	$13.64^{+0.99}_{-0.91}$	$(12.67^{+0.75}_{-0.69})$	(1)	(0.2)	$0.79^{+0.15}_{-0.25}$	< 0.2	$8.28^{+0.23}_{-0.32}$	1.3
DC free	-21.125	$11.35^{+0.13}_{-0.05}$	$(12.12^{+0.16}_{-0.05})$	$(11.51^{+0.12}_{-0.04})$	(1)	(0.2)	$0.36^{+0.3}_{-0.12}$	$3.87^{+0.24}_{-0.64}$	$7.65^{+0.57}_{-0.18}$	1.5
DC = 0.6	-21.125	$11.48^{+0.02}_{-0.02}$	$(12.26^{+0.03}_{-0.03})$	$(11.62^{+0.02}_{-0.02})$	(1)	(0.2)	(0.6)	$3.29^{+0.11}_{-0.1}$	$8.16^{+0.1}_{-0.1}$	1.3
DC = 1	-21.125	$11.51^{+0.02}_{-0.02}$	$(12.3^{+0.03}_{-0.02})$	$(11.65^{+0.02}_{-0.02})$	(1)	(0.2)	(1)	$3.16^{+0.08}_{-0.1}$	$8.3^{+0.12}_{-0.08}$	1.9
Bias Only	-21.125	NA	NA	NA	NA	NA	NA	NA	$10.86^{+0.1}_{-0.2}$	0.8
Harikane16	-20.0	$11.30^{+0.10}_{-0.13}$	$(12.06^{+0.07}_{-0.16})$	$(11.47^{+0.05}_{-0.12})$	(1)	(0.2)	(0.6)	5.0	$6.3^{+0.4}_{-0.4}$	0.5
Harikane16	-19.1	$11.03^{+0.05}_{-0.18}$	$(11.75^{+0.20}_{-0.29})$	$(11.23^{+0.15}_{-0.22})$	(1)	(0.2)	(0.6)	7.1	$5.5^{+0.2}_{-0.4}$	1.4

Table 5.2: Comparison of abundance matching results to clustering fits for the $z \sim 6$ LBGs. Columns are (1) LBG sample used, (2) LBG threshold absolute magnitude, (3), observed comoving number density (Mpc^{-3}), (4) the minimum halo mass in the most straightforwards abundance matching scheme (log base ten Solar mass units), (5) the model comoving number density (Mpc^{-3}) of the best fit HOD models in this work and Harikane et al., (2016) without incorporating duty cycle, (6) the corresponding minimum halo mass from the HOD model (log base ten Solar mass units).

Data	M_{UV}	n_g^{observed} (Mpc^{-3})	$\log M_{min}^{\mathrm{matched}}$	n_g^{model} (Mpc^{-3})	$\log M_{min}^{\mathrm{model}}$
Bowler15	-21.125	4.1×10^{-5}	11.51	8.9×10^{-6}	$11.53^{+0.05}_{-0.07}$
Harikane16	-20.0	3.8×10^{-4}	11.09	2.1×10^{-4}	$11.30^{+0.10}_{-0.13}$
Harikane16	-19.1	13.4×10^{-4}	10.79	7.3×10^{-4}	$11.03^{+0.05}_{-0.18}$

5.5 Discussion

5.5.1 The link between low- and high-luminosity galaxies and their haloes at $z \sim 6$

The most relevant study to compare the results presented here on the bright LBGs is with Harikane et al. (2016), who presented clustering measurements and HOD fits to $z \sim 6$ LBG galaxies, but on smaller angular scales of approximately $10^{-3.25} < \theta/\mathrm{deg} < 10^{-1.25}$, compared with $10^{-3} < \theta/\mathrm{deg} < 10^{-0.5}$ here, and for fainter rest frame absolute magnitudes of $-20.5 < M_{UV} < -19$ compared with $-22.7 < M_{UV} < -21.125$ here (see Fig. 5.6). Thus my results combined with Harikane et al. (2016) describe LBG clustering over almost three orders of angular scale and a factor of 40 in luminosity.

The bias and halo mass results compared with the results of Harikane et al. (2016) are shown in Fig. 5.7. Although only moderate quality fits (possible reasons for which are discussed in the subsequent sub-sections), all fitted models suggest the bright LBG sample has a substantially higher typical host halo mass and galaxy bias than the lower luminosity samples in Harikane et al. (2016). This higher bias is evident by directly comparing the two measurements of the correlation function. The bright LBG sample has an amplitude ~ 3 times higher than the Harikane et al. (2016) bright ($M_{UV} < -20$) sample with $\omega(0.01^\circ) \sim 0.2$, and the measured bias is a factor of 1.7 greater (as $\omega \propto b^2$). In general higher luminosity and higher stellar mass galaxy samples have higher biases, but it is important to note that

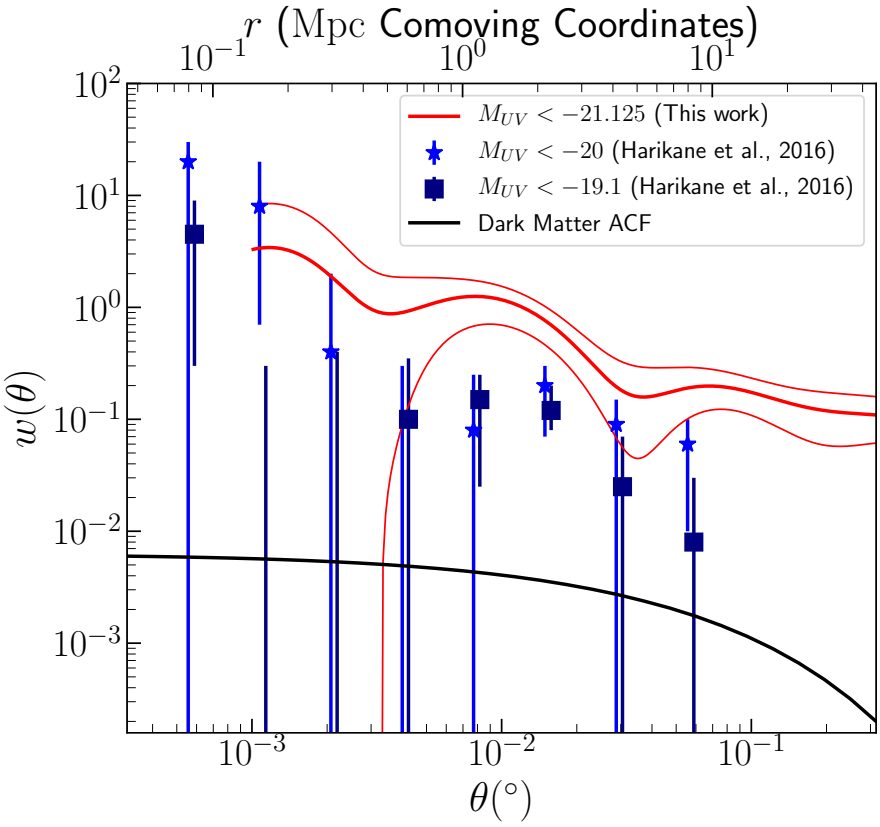


Figure 5.6: Comparison of the LBG correlation function (red curve, 1-σ uncertainties in the lighter curves), with the lower luminosity Harikane et al., (2016) measurements, and the dark matter angular correlation function (black curve).

it was not a foregone conclusion to measure a bias this high. It was entirely possible that the $M_{UV} \sim -21.5$ sample could have been the same (or largely the same) population as the sample of Harikane et al. (2016), just observed during a particularly vigorous but rare burst of star formation. If that had been the case, the clustering amplitude would have been lower, and the inferred duty cycle much lower. The comoving space density of the bright LBGs used in this chapter is 4.1×10^{-5} Mpc^{-3}, compared with 3.8×10^{-4} Mpc^{-3} for the most luminous $z = 6$ galaxies in Harikane et al. (2016). Harikane et al. (2016) do not measure the duty cycle for this sample, but assume it to be equal to 0.6. As an illustrative example, a duty cycle of 0.6 for the Harikane et al. (2016) sample would mean an actual underlying

population comoving density of $6.3 \times 10^{-4}\mathrm{Mpc}^{-3}$. If the bright LBG sample investigated in this Chapter was part of the same population, that would correspond to DC = 0.06 (in other words that the fainter population spends approximately 6 percent of its time in this super-enhanced state of star formation). However the amplitude of the clustering rules this out and the $M_{UV} < -21.125$ sample is comprised of continuously high-luminosity objects in very dense environments.

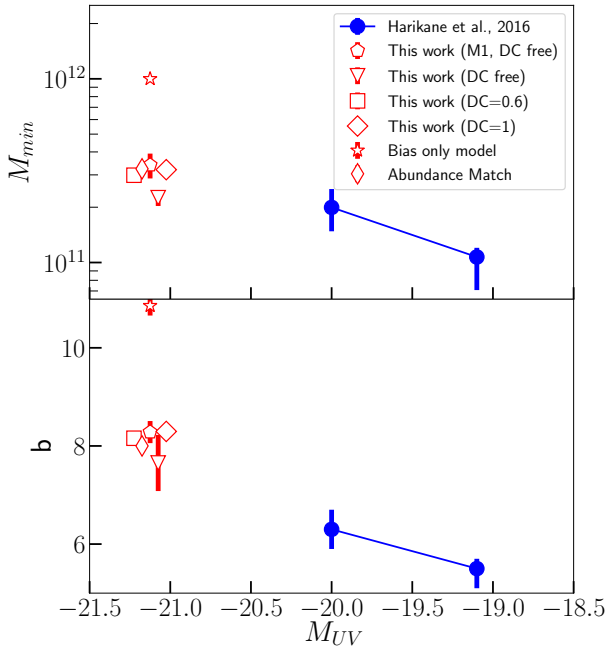


Figure 5.7: Comparison of the results from the LBG sample in this chapter, with comparable measurements of lower luminosity LBGs from Harikane et al. (2016). Top: M_{min} as a function of absolute UV luminosity threshold (in units of Solar mass). Bottom: galaxy bias as a function of absolute UV luminosity threshold. The results from the six different models are shown for comparison (x-axis values slightly offset for each model for clarity).

5.5.2 Apparent lack of a 1-Halo term

Models using extrapolated values of M_1 suggest that at scales of $10^{-2.5}$ deg and smaller there should be a sharp upturn in the value of the correlation function as the observations start to probe clustering of multiple galaxies within individual halos (see Fig. 5.4). This is not apparent in the bright LBG sample, in contrast to Harikane et al. (2016) (see Fig. 5.6). A direct interpretation of this would be that M_1 just increases much faster than the extrapolation of Equation 5.1 i.e. the satellite fraction drops off extremely fast and an

unfeasibly large (for this redshift) halo is needed to host two of these sources. Another possibility is suggested by Jose et al. (2013), who also observe a lack of a 1-halo term in clustering measurements of $z \sim 3-7$ LAEs. Their proposed solution was that halo occupancy behaved in a sub-Poissonian manner, and they found that a modified distribution (see their equation 15) was able to reproduce the measurements. However, I suggest that there are good reasons to believe that there is strong cosmic variance on the small scale measurements that is not accounted for in the bootstrap uncertainties, making it difficult to make direct inferences about the satellite population of these galaxies.

For a single contiguous field observation, cosmic variance is smaller on small scales than on large scales in a limited sense, simply because one observes more instances of small scale structure. However these are not independent instances of small-scale structure, as they all come from the same large-scale density field. To illustrate this, suppose the sources have $M_{\rm min} \sim 10^{11.4} M_{\odot}$, then one would expect $M_1 \sim 10^{12.2} M_{\odot}$ e.g. only halos with $M > 10^{12.2} M_{\odot}$ host more than one of the bright sample. The comoving density of $M > 10^{12.2} M_{\odot}$ halos is $5.2 \times 10^{-7} {\rm Mpc}^{-3}$ and the comoving volume probed by the observations is $\sim 1.7 \times 10^{7} {\rm Mpc}^{3}$. This means that the expected numbers of $M > 10^{12.2} M_{\odot}$ halos in the volume surveyed is ~ 10. Just these ten would give ~ 10 close pairs (in addition a few more would be expected from projection effects), which is more than twice the four close ($10^{-3} - 10^{-2.5}$deg) pairs observed here, and would push the small scale correlation function to higher values. However these halos will be extremely biased, much more so than the $\sim 10^{11.4} M_{\odot}$ halos. Conceivably for an extreme case, it could be that if the observations were repeated 10 times, one would find that 9 times no $M > 10^{12.2} M_{\odot}$ haloes were observed, and the tenth time a very overdense region is observed, which contains 100 haloes in. In the first nine cases, no satellites would be observed, leading to a flat correlation function to small radii that is seen in Fig. 5.2, and the tenth time an overestimate of the satellite fraction is measured. Therefore one would expect there to be very substantial cosmic variance on the measurements of the correlation function on small scales - variance that is not incorporated into the errors on clustering measurements. Essentially the 1-halo term is dominated by contributions from very massive halos, which are the most biased, so there is the most cosmic variance on small scale measurements of the correlation function. It appears to be

the case that neither of the fields are overdense enough to sample the highly biased sample of massive halos at this redshift which could be massive enough to host multiple bright LBGs, and instead the small scale measurements are dominated by the angular projection of the linear clustering (e.g. objects near in angular space by chance, but not near in physical space).

The Harikane et al. (2016) measurements however do have a prominent 1-halo term. I suggest that the reason for this may lie in the fact that a) they are at lower luminosities, so the cosmic variance on halos required to host multiple galaxies is less extreme and b) their correlation functions are measured from galaxies in seven different fields (rather than my two), so they had a greater chance of observing a dense field that had the massive halos necessary for satellites.

5.5.3 Mismatch between the number counts and bias measurements

As discussed in Section 5.4.2, it was not possible to obtain a good HOD fit to the number counts and clustering. A good fit was only obtained for the clustering measurements with a pure bias model. The core of the discrepancy is that to obtain the directly measured bias of ~ 10, the galaxies would need to be in halos of minimum mass $M_{\rm min} \sim 10^{12} M_{\odot}$. This corresponds to a comoving density of $2.8 \times 10^{-6}{\rm Mpc}^{-3}$, compared with an observed number density of $4.1 \times 10^{-5}{\rm Mpc}^{-3}$ i.e. approximately a factor of 15 lower than the observed value. Similarly, plain abundance matching would suggest a minimum mass of $M_{\rm min} \sim 10^{11.5} M_{\odot}$ corresponding to a bias of ~ 8. I note that Barone-Nugent et al. (2014) report a very similar issue at $z \sim 7.2$, where they found that a duty cycle of 1 was needed for their LBG sample, and that even then the measured bias was slightly inconsistent with the number density. I discuss in this sub-section possible explanations for this discrepancy.

5.5.3.1 Contaminants

It is possible that some of the sources used for the clustering measurements are not truly $z \sim 6$ LBGs, but are instead brown dwarfs or galaxies at other redshifts. This is unlikely to be the cause of the discrepancy, as Bowler et al. (2015) had access to photometry across a very large range of wavelengths and performed extensive testing with brown dwarf templates

to rule out substantial contamination. Furthermore, stellar contamination would actually reduce the clustering amplitude as stars are unclustered and have no physical correlation with the galaxies. This is also in general the case for contamination from galaxies at other redshifts, which would not be spatially correlated with the region of space probed at $z \sim 6$. One possible exception to this is if a substantial proportion of the sources are actually at $z \sim 1.3$ (the redshift degenerate with $z \sim 6$ when it is difficult to distinguish between the two spectral breaks in SED modelling), in which case I would effectively be measuring the correlation function at $z \sim 1.3$, which would render all modelling so far void. However I dismiss this possibility - although it is possible there are a few $z \sim 1.3$ interlopers in the sample, it seems extremely unlikely that they make up a substantial proportion of the sample a) because of confidence in the template fitting, and b) because of the deep multi-wavelength data used in Bowler et al. (2015). I conclude that contamination is unlikely to be a substantial factor in the clustering-number density discrepancy.

5.5.3.2 More complex galaxy-halo relations

HOD modelling is predicated on the principle that the only thing that determines the galaxy content of a halo is the mass of the halo - if this is violated, then in general more complex relations between the galaxy-halo relation and clustering measurements are possible. Well known cases include assembly bias (see Hearin et al., 2016a, where the bias of halos depends on halo assembly history as well as mass), a requirement of interactions with a nearby satellite halo (Cen & Safarzadeh, 2014), or a dependence on the large scale linear density field. These effects are all plausible - it is possible to imagine a galaxy having a brief starburst (which is then observed as an ultra-bright source) as a result of an interaction in a denser region of the Universe, or for the amount of gas left for star formation in a halo to be related to the age of the halo. However it is in general very hard to distinguish between these different effects if one only has access to the large-scale linear bias (which is effectively a one dimensional measurement). Galaxy-galaxy lensing can in principle observationally break these degeneracies (e.g. if sources are in older, lower mass halos, their clustering will reflect the assembly-dependent bias of the hosts, but the lensing will reflect just the mass), but this is likely to never be possible at these redshifts as it requires a high number density

of even higher redshift sources. It seems likely that comparison with simulations is the only way to investigate the viability of such underlying processes.

5.5.3.3 Uncertainty in knowledge of the high-redshift dark matter distribution

A key input to HOD modelling is knowledge of the spatial distribution of the underlying dark matter, in particular the HMF and halo bias as a function of halo mass (which comes predominantly from N-body simulations). Fitting HMF and halo bias models at high redshift remains a challenging endeavour. If the dark matter model used is incorrect, then the conclusions from HOD modelling will also in general be incorrect. As shown in Table 5.1, abundance matching suggests that the sources are in $> 10^{11.5} M_{\odot}$ halos. However, the bias from the clustering would suggest that they are in $> 10^{12} M_{\odot}$ halos, which are a factor of twenty rarer. Tinker et al. (2008) and Tinker et al. (2010) found model HMFs and halo biases from N-body simulations at redshifts of $z = 0 - 2.5$. Behroozi et al. (2013) then introduced a high-redshift calibration to the Tinker et al. (2010) HMF, extending the validity to $z \sim 8$ (representing an increase of approximately 20 percent at $z = 6$ for $M \sim 10^{11.3} M_{\odot}$ halos). Hence confidence in N-body simulations makes a correction factor of ~ 20 appears implausible, within the current structure formation paradigm. However, Behroozi et al. (2013) did not calibrate the high redshift halo bias, so I am effectively using biases at $z \sim 2.5$ extrapolated to $z \sim 6$. The excess in the clustering amplitude is only around 50 percent, which would require around a 25 percent correction in bias. Thus I suggest that the results presented in this chapter could potentially be explained with a high-redshift calibration to the halo bias function that steepens it at the high mass end. Behroozi & Silk (2016) also discuss this direction of inference e.g. how high redshift stellar-mass functions can give information on the high-redshift HMF.

An alternate potential correction to our understanding of the distribution of dark matter is the incorporation of 'quasi-linear effects'. HOD modelling makes a binary division between non-linear clustering within halos, and large-scale linear bias. However this transition is gradual, not sharp, and bias can be scale-dependent on up to 10 Mpc scales (relevant for scales probed with these observations), although it always tends to a constant value at large scales (Mann et al., 1998). Introducing a functional form for scale-dependent bias can

model some of these effects and Jose et al. (2016) conclude that quasi-linear clustering has the largest effect at high redshift ($z > 2$), and high halo mass. In particular, Jose et al. (2017) note a similar discrepancy in the number counts and clustering at $3 < z < 5$, and find that quasi-linear effects can cause one to over estimate halo mass by up to a factor of ten if unaccounted for. They give γ values (the correction to the bias) of order 30-40 percent - around the size of the inconsistency - at the relevant masses and scales relevant for this analysis, so it seems plausible that incorporating quasi-linear effects could solve the discrepancy, and be necessary for future analysis. Jose et al. (2017) also show that quasi-linear effects make the transition into the 1-halo term less sharp, which could explain why I do not see evidence for the 1-halo term in these measurements, as discussed in section 5.5.2. Quasi-linear effects are sub-percent at lower redshift, so HOD modelling at lower redshift is not invalidated (van den Bosch et al., 2013).

5.5.3.4 Modification of our understanding of high redshift structure formation

Alternatively, it may be the case that N-body simulations do not correctly capture the physics of early structure formation, in a way that no calibration will be able to account for. The potential issue of 'too many' high mass/luminosity galaxies has been identified by Steinhardt et al. (2016). They summarise results that suggest that the best constraints on the HMF at $z = 4-10$ inferred from observations is dramatically higher than the HMF from ΛCDM, with the discrepancy getting worse towards high halo masses and higher redshifts. Although in a challenging observational regime, they suggest that the observations show that current theories of structure assembly at $z > 4$ could be flawed.

Although a possibility, the results in this chapter are not in sufficient disagreement with models to warrant support of this hypothesis yet - it is necessary to explore the much more likely possibilities of the high-redshift halo bias needing calibration and quasi-linear issues before considering more dramatic changes to theories of structure formation.

5.5.4 Estimating Cosmic Variance

Cosmic variance is a term that can be used to refer to a number of related but subtly different effects. The specific context in which I use the term here is that many extragalactic

statistical measurements vary by more than sample variance between different fields because of large scale structure. As noted in Bowler et al. (2015), the number density of the two fields varies by much more than sample variance assuming a Poisson distribution. This is a consequence of large-scale structure, which clustering measurements quantify. These clustering measurements can be linked back to the number count estimates to see if the cosmic variance observed is consistent with the clustering measurements, or if one of the fields is over/under dense, even accounting for large-scale structure. Understanding cosmic variance can be important for correctly connecting high redshift observations of galaxies with our understanding of reionization e.g. Ouchi et al. (2009).

Note that in general it is possible for two populations to have the same average number counts, but different cosmic variances - this occurs when they have the same 1-point statistics, but different 2-point statistics. Thus the 2-point statistics can be used to refine the estimate of cosmic variance in Bowler et al. (2015) who used the Trenti Cosmic Variance calculator (Trenti & Stiavelli, 2008) - which only uses 1-point statistics. A clustered and unclustered population of the same number density will have substantial and zero cosmic variance respectively (both will have Poisson variance). Also 2-point statistics only give the *variance* of the full probability distribution of counts in a field. Higher order statistics (n-point correlation function etc.) are needed to probe the full distribution. Similarly, 3-point statistics are needed to quantify the cosmic variance on measurements of 2-point statistics, 4-point statistics are needed to quantify the cosmic variance on measurements of 3-point statistics, and so on ad infinitum. Alternatively, more complex cosmic variance behaviours can be studied with mock catalogs from cosmological simulations of structure formation (Trenti & Stiavelli, 2008).

The cosmic variance is related to the *expected value* of the correlation function in the geometry of the field, that is to say the expected value of the correlation function at the separation of two points randomly selected in the field. Analytically the expectation can be written as:

$$\bar{\omega}(A) = \frac{\int_A \int_A \omega(|\vec{\theta_i} - \vec{\theta j}|) d^2\theta_i d^2\theta_j}{\int_A \int_A d^2\theta_i d^2\theta_j}, \tag{5.6}$$

where A is the angular region of the field, ω is the 2-point correlation function, $\bar{\omega}(A)$ is the expectation of the correlation function in that field, θ_i and θ_j are points in the field, $|\vec{\theta_i} - \vec{\theta j}|$ is their angular separation, and the integrals are double integrals over the area of the field. I calculate this numerically by sampling 100,000 pairs of points in the field, calculating their angular separation, finding the value of the correlation function at that angular scale (with the best fit model from Section 5.4.2), and then taking the average.

Trenti & Stiavelli (2008) summarise results from Peebles (1993); Newman & Davis (2002) and Somerville et al. (2004) that conclude:

$$\bar{\omega}(A) = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle^2} - \frac{1}{\langle N \rangle}, \tag{5.7}$$

where N is the random variable of the number of objects in a field. This can be rearranged to the form:

$$\langle N^2 \rangle - \langle N \rangle^2 = \langle N \rangle + \bar{\omega}(A) \langle N \rangle^2, \tag{5.8}$$

and,

$$\sigma^2_{total} = \sigma^2_{poisson} + \sigma^2_{CV},$$

where $\sigma_{total} = \sqrt{\langle N^2 \rangle - \langle N \rangle^2}$ is the total standard deviation on measurements of number counts, $\sigma_{Poisson} = \sqrt{\langle N \rangle}$ is the Poisson standard deviation and $\sigma_{CV} = \sqrt{\bar{\omega}(A)} \langle N \rangle$ is the standard deviation from cosmic variance e.g. total standard deviation is the Poisson and cosmic variance standard deviations added in quadrature. The standard deviation from cosmic variance reduces to $\sigma_{CV} = b\sqrt{\bar{\omega}_{DM}(A)} \langle N \rangle$ in the 'pure-bias' case where $\omega = b^2 \omega_{DM}$, and b is the bias and ω_{DM} is the dark matter angular correlation function.

This formalism has all the properties one would expect from cosmic variance. Cosmic variance is higher when sources are more clustered (further away from uniform). Cosmic variance becomes lower as the size of the field increases. This is because the correlation function is sampling larger scales, where the function has a lower value, a consequence of the fact that a larger range of environments is being probed. A more subtle effect is

Table 5.3: Actual and expected number of galaxies in each field. The columns are: field used, the field angular area (in deg^2), the actual number of galaxies in the field ($N_{\rm a}$), the actual angular galaxy density in the field ($\rho_{\rm a}$, in deg^{-2}), the expected number of galaxies in the field if it had the mean density ($N_{\rm e}$), the expected angular galaxy density in the field - all identical figures as I am considering deviations from the mean density ($\rho_{\rm e}$, in deg^{-2}), standard deviation from Poisson statistics, equal to the square root of $N_{\rm e}$ ($\sigma_{Poisson}$), standard deviation from cosmic variance, estimated from the clustering measurements (σ_{CV}), and the Poisson and cosmic variance errors added in quadrature (σ_{Total}).

Field	Area (deg^2)	$N_{\rm a}$	$\rho_{\rm a}$ (deg^{-2})	$N_{\rm e}$	$\rho_{\rm e}$ (deg^{-2})	$\bar{\omega}$	$\sigma_{Poisson}$	σ_{CV}	σ_{Total}
UDS	0.74	64	86	92	124	0.027	10	15	18
UltraVISTA	0.62	103	166	77	124	0.023	9	12	15
Total	1.35	167	124	167	124	0.013	13	19	23

that cosmic variance also varies with field shape, as well as size. The average length scale probed for a circle is a lot smaller than for a long thin rectangle of the same area for example, corresponding to a higher average correlation function value, and greater cosmic variance. This can be interpreted (Trenti & Stiavelli, 2008) as a consequence of the fact that a more compact field geometry is predominantly sampling the same environment, be it an over- or under-density. However a long thin geometry is sampling from a large range of environments, and overdensities and underdensities are more likely to cancel out. The formalism for describing cosmic variance here also works when the field is disconnected. If the 'field' is actually two disconnected subfields separated by a vast distance in the sky, when calculating the average of the correlation function over this field, half the time the two points will be in different sub-fields, and the value of the correlation function on this scale will be effectively zero. This halves the value of $\bar{\omega}(A)$, effectively reducing cosmic variance contribution by a factor of $\sqrt{2}$ as completely different regions of the Universe are being probed.

Table 5.3 summarises my results relating to cosmic variance when applied to the UltraVISTA and UDS fields for the $z \sim 6$ samples investigated in this chapter (using the best fit pure bias model). The most important columns to compare are the N_a (the actual number of galaxies in the field), N_e (the expected number there would be if both fields had the average density) and σ_{total} the standard deviation on counts including Poisson and cosmic variance implied by my clustering measurements. In both cases, the observed number of galaxies in each field is approximately a 1.5σ deviation from the model value (as noted in Bowler et al., 2014, they are the most over- and under-dense respectively of the five CAN-

DELS fields). Thus at these redshifts, both UltraVISTA and UDS appear to be moderate, but not unreasonable, over and under-densities respectively.

5.5.5 The onset of quenching

Bowler et al. (2014) and Bowler et al. (2015) report a rapid evolution in the high luminosity end of the luminosity function at $z = 6 - 7$, a transition from a power-law drop off to an exponential cut-off, interpreted as the onset of quenching or dust obscuration. I show in Fig. 5.8 the luminosity to halo mass ratios implied by the Bowler et al. (2015) UV-luminosity functions and the simplest possible abundance matching scheme:

$$M_h(M_{UV}) = \psi^{-1}_{HMF}(\psi_{lum}(M_{UV})),$$

where M_h is the halo mass corresponding to the magnitude M_{UV}, $\psi_{HMF}(M_h)$ is the comoving number density of halos greater than mass M_h and $\psi_{lum}(M_h)$ is the comoving number density of LBGs brighter than magnitude M_{UV}. This is a very simple model (more complex abundance matching schemes do exists e.g. SHAM, Guo et al., 2016), and ignores complexities such as scatter, satellites and duty cycle, but illustrates the argument in Bowler et al. (2014) that $z = 6 - 7$ is the onset of quenching. At the low luminosity end, the luminosity to halo mass ratio drops with time, as Harikane et al. (2016) find (see their figure 10). At the high mass end, the ratio is fairly constant, before dropping off towards low redshift (e.g. Behroozi et al., 2013). The unphysical rise in the $z = 7$ luminosity to halo mass ratio at the high luminosity end is a result of the fact that the luminosity function cannot (as noted in Bowler et al., 2014) continue as a power law to even brighter magnitudes, as it would quickly be dramatically higher than the HMF. More realistically, in a scenario with no high luminosity/mass end quenching, the luminosity function would drop off at the same rate as the HMF (which does not drop off as fast as a Schechter function). If at higher luminosities it really does continue as a power law, then the LBGs would become much more numerous than their corresponding halos, and almost certainly these bright objects would be rare phases in the duty cycle of a more common population. Davidzon et al. (2017) make a similar argument for the onset of quenching, using stellar mass estimates from

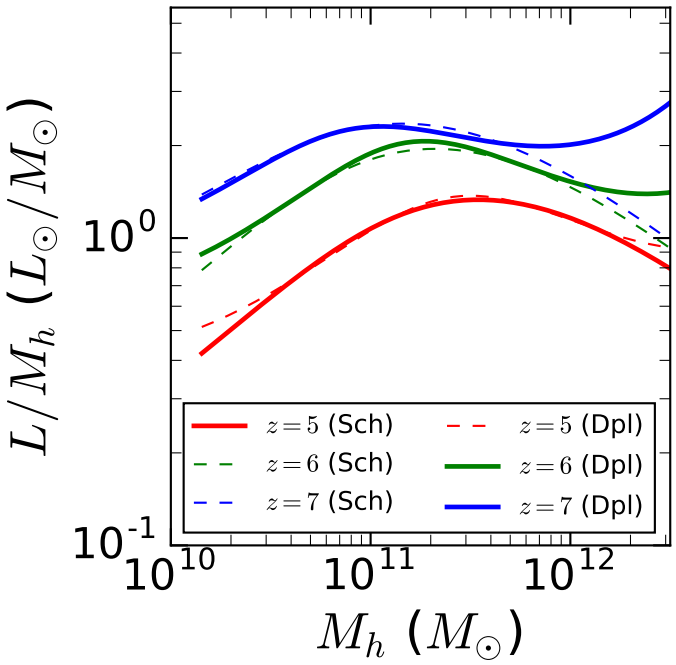


Figure 5.8: Luminosity (monochromatic luminosity at 1500$\mathring{A}$; that is to say, λf_λ where λ is the wavelength and f_λ is the flux per wavelength) to halo mass ratios as a function of halo mass, derived using abundance matching with the Bowler et al., (2015) luminosity functions and Behroozi et al., (2013) halo mass functions. The results are shown for both Schechter function fit and the double power law fit, for $z = 5$, $z = 6$ and $z = 7$. In bold are the best fitting functions, double power law for $z = 6$ and $z = 7$, Schechter for $z = 5$, and the dashed line shows the alternative.

UltraVISTA-DR2, SPLASH and Hyper-Suprime-Cam data in the COSMOS field, except finding the transition at $z \sim 3$ rather than $z \sim 6$.

The fact that these galaxies at the bright end of the luminosity function truly are in the densest regions of the Universe as opposed to less biased objects caught in extremely rare massive star-bursts, supports the explanation that the drop off in the luminosity to halo mass ratio at the high-mass end is essentially still operational at $z = 6$. If the sample had been just rare episodes of vigorous star formation, then the interpretation of the steepness/drop off rate at the bright end of the luminosity function would be different - it would instead be dominated by the distribution of star-formation rates a given population has, and how rare its episodes of high star formation are, as opposed to the luminosity function being dominated by a modified halo mass function.

To determine if $z \sim 6-7$ is the onset of mass quenching, a similar analysis to this work of luminous $z = 7$ LBGs would need to be performed to see if the luminosity to halo mass ratio does not drop off. There are three main possibilities at $z \sim 7$:

- $M_{UV} \sim -22$ LBGs have the same host halo mass as $M_{UV} \sim -20$ objects; this would suggest they are the same population of objects at different points in their duty cycle
- $M_{UV} \sim -22$ LBGs have the same host halo mass as they do at $z \sim 6$. This would suggest that the luminosity to halo mass ratio at the high luminosity end does not change much over $z \sim 6-7$, which would not support $z \sim 6-7$ being the onset of mass quenching/dust obscuration
- $M_{UV} \sim -22$ LBGs have a lower host halo mass as they do at $z \sim 6$, but still higher than the galaxies with $M_{UV} \sim -20$, in such a way that the luminosity to halo mass ratio was constant as a function of halo mass. This would be supportive that $z \sim 6-7$ was indeed the onset of quenching (or dust obscuration).

5.6 Conclusions

I have used the largest existing sample of extremely bright Lyman-break galaxies at $z \sim 6$ to investigate their large scale structure and links to the possible onset of feedback quenching or dust obscuration at this redshift. This sample (detailed in Bowler et al., 2015) of 263 LBGs was selected in the UltraVISTA/COSMOS and UDS/SXDS fields, using deep optical and near-infrared data required to distinguish the galaxies from contaminant populations. The method I used to study the connection between the galaxies and their host halo was to measure their clustering with the angular correlation function, and model these measurements with a HOD scheme.

The key conclusions of this work are:

- Bright LBGs ($M_{UV} \leqslant -21$) appear to be highly biased ($b \sim 10$) objects in dense environments, as opposed to being rare temporal episodic incarnations of fainter galaxies ($M_{UV} \sim -19$). This suggests that the bright-end of the luminosity function at $z \sim 6$ is determined by feedback processes or dust obscuration, rather than duty cycles. My

results have important implications for the physical origin of the observed steepening of the bright end of the ultra-violet luminosity function between $z \sim 6$ and $z \sim 7$ (Bowler et al., 2014, 2015) - which in a straightforward abundance matching scheme would imply a dramatically increased luminosity to halo mass ratio at $z \sim 7$ to $z \sim 6$.

- I find a tension between the observed number counts and bias, that suggests that some modification to our knowledge of the high-redshift dark matter distribution is needed This is most likely to be the incorporation of quasi-linear effects (as described in Jose et al., 2017), or possibly a minor calibration upwards of halo bias at high redshift.

- Although number counts within each field differ by far more than Poisson sample variance, estimates of the cosmic variance from the clustering would suggest that both fields are only moderate 1.5-σ over/under densities.

- A low duty cycle is not required to explain the observations (equivalently DC ~ 1), and the satellite fraction of the sources is very small, at most a few percent

Chapter 6

Simulating the Universe

The work in this Chapter compares the observations from Chapter 3 to hydrodynamic cosmological simulations, and will be submitted to a journal in the coming months.

6.1 Introduction

Parallel to the great success of wide-field surveys in *observing* large numbers of galaxies, there have also been great advances in *simulating* large numbers of galaxies in cosmological simulations (see Somerville & Davé, 2015 for a review). Cosmological simulations typically model the Universe as a finite cube with periodic boundary conditions, take some set of initial conditions, some formulation of the physics of the Universe that the simulator is interested in capturing, and 'run' the universe from the beginning of time up to $z = 0$. Critically, simulations are the only way to solve the full growth of cosmological structure into the strongly non-linear regime (including baryonic physics) - otherwise we are limited to comprehending only dark matter in the linear regime (see Chapter 1). Cosmological simulations have dramatically changed over the last $\sim$ 50 years as computing power has improved, from being only able to qualitatively capture the large-scale structure of the Universe, to being able to capture in great detail a wide range of baryonic and galaxy physics.

There are three main categories of cosmological simulations:

- *'N-body'* simulations, as discussed in section 1.2.2.4, model dark matter only, allowing

'N' dark matter particles to evolve under gravity. N-body simulations are an important probe of the large-scale structure of the Universe. However they contain no galaxy or baryonic physics. Contemporary N-body cosmological simulations can have $N \sim 10^{11}$, probing tens of cubic Gigaparsecs, sensitive to halos of mass $10^{11} M_{\odot}$ and above e.g. Kim et al. (2011).

- *Semi-analytic models* (SAMs) take a pre-existing N-body simulation, and 'graft' a galaxy population on top according to some prescription. Early forerunners of SAMs populated the haloes as per HOD e.g. simply place galaxies in haloes according to some function of halo mass, see Benson et al. (1999); Berlind & Weinberg (2002) - but this approach contains no physical processes. Typical contemporary SAMs model galaxy growth over cosmic time consistently with the dark matter halo merger tree, and typically also track a large range of physical parameters and processes e.g. total gas accreted into a halo, gas temperature, removal of cold gas from the galaxy due to feedback, how mergers transform disks into spheroids (Benson, 2010; Somerville & Davé, 2015). The model that describes how the baryons and galaxies behave within the simulation is typically characterised by a few parameters, which can then be tuned to observations e.g. Lu et al. (2011). Notable examples of SAMs include Galform (Bower et al., 2006), Semi-Analytic Galaxy Evolution (SAGE, Croton et al., 2006, Croton et al., 2016), SeSAM (Neistein & Weinmann, 2010) and the model of Guo et al. (2012) in the Millennium simulation (Springel et al., 2005).

- *Hydrodynamic* simulations seek to incorporate even more physics in a more fundamental way, directly simulating gravity, hydrodynamics, and thermodynamics simultaneously (normally using either particles, grid cells, or a combination). Hydrodynamic simulations typically model dark matter, gas and stars simultaneously, incorporating physical processes like feedback and star formation (see below). Notable examples include Horizon-AGN (Dubois et al., 2014), the Evolution and Assembly of GaLaxies and their Environments project (EAGLE, Schaye et al., 2015; McAlpine et al., 2016), Illustris (Vogelsberger et al., 2014), Mufasa (Davé et al., 2016) and MassiveBlack-II (Khandai et al., 2015).

Hydrodynamical simulations are typically the most successful at recreating galaxy population properties self-consistently with comparatively few tuneable parameters (at the cost of being vastly more computationally expensive than SAMs). Physics that most hydrodynamical simulations are designed to capture are gravity, hot gas hydrodynamics, star formation, super-massive black hole formation and evolution, AGN feedback, stellar feedback and chemical evolution of stellar populations. Hydrodynamic simulations have made great contributions in recent years to many areas of galaxy evolution and cosmology, including understanding which galaxies are responsible for reionization (Sharma et al., 2016), the size evolution of compact galaxies (Furlong et al., 2015), the alignment of galaxies with the cosmic web (Dubois et al., 2014; Chisari et al., 2015), and the role of AGN (Beckmann et al., 2017). In this Chapter I focus on the Horizon-AGN simulation, a hydrodynamic simulation with up to date star formation, gas cooling and AGN feedback prescriptions, see (Dubois et al., 2014).

No simulation can capture all physics present in the real Universe. Instead we wish to investigate how successful a simulation is in recreating observations and understand which aspects of galaxy formation physics it correctly captures. One popular approach is to use 1-point statistics as a measure of how successful a simulation is e.g. how well does the simulation recreate the observed stellar mass function (see Schaye et al., 2015). Similarly star formation rates (Katsianis et al., 2016) and morphology (Dubois et al., 2016) can be compared between observations and simulations. However there is comparatively little work comparing the large-scale spatial distribution/clustering of galaxies in observations to galaxy clustering in hydrodynamic simulations. 1-point statistics are not sufficient to assess all aspects of the quality of the simulation - the spatial distribution of the galaxies gives extra information e.g. on the scale at which baryonic processes are important and how galaxies interact which each other and their environment. Early work in measuring and understanding galaxy bias in simulations was carried out by Blanton et al. (1999, 2000). Building on this, most clustering analyses comparing observations to hydrodynamic simulations have focussed on galaxy bias (e.g. Cen & Ostriker, 2000, Weinberg et al., 2004, Yoshikawa et al., 2001). However the bias does not include all information about the spatial distribution of galaxies as this does not include smaller scale non-linear processes. There is

more work in the literature comparing clustering from SAMs to observations. Li et al. (2012) compare SDSS and DEEP2 clustering from the Guo et al. (2012) SAM. Saghiha et al. (2017) compare galaxy-galaxy lensing (a closely related statistic to clustering) from CHFTLens to corresponding lensing predictions from a Millennium SAM. The comparison analyses in the literature most relevant for this chapter are McCracken et al. (2007) and Artale et al. (2016). McCracken et al. (2007) (using a SAM), compare the angular correlation function of galaxies over a range of luminosities in CHFT observations and Millennium mock catalogues at $z \sim 0.8$ on scales $0.001^\circ < \theta < 0.1^\circ$, probing both linear and non-linear scales. Artale et al. (2016) compare clustering in the EAGLE hydrodynamic simulation to observational clustering results from the GAMA survey (Driver et al., 2011) at $z \sim 0.1$, finding good agreement on non-linear scales.

In this chapter I seek to build on previous work comparing observations and simulations, by comparing clustering and HOD measurements (as a function of stellar mass) between observations and hydrodynamic cosmological simulations. In particular I compare clustering measurements and HOD modelling between VIDEO observations and mock catalogues from the Horizon-AGN hydrodynamic simulation, as they are both sensitive to similar stellar masses, length scales and redshift ranges for a key part of the history of the Universe.

This Chapter is organised as follows: first I briefly describe the Horizon-AGN simulation, and the mock catalogue constructed from it (section 6.2). I then measure the correlation function in the simulation mock catalogue for sub-samples corresponding to those considered in VIDEO in Chapter 3 (section 6.3.1). I fit HOD models to these simulated observations, and compare and contrast the Horizon-AGN results to those from the VIDEO observations (section 6.3.3). Finally I discuss the implications of my results for understanding which aspects of galaxy formation physics simulations capture well, and which they don't, and for understanding how confident we can be in inferences from HOD modelling (section 6.4).

6.2 Simulation Data

In this section I discuss the Horizon-AGN simulation used in this Chapter, and the mock skies constructed from the simulation that I compare with the observations.

6.2.1 Horizon-AGN

The full specifications of Horizon-AGN and the physics that is incorporated into the simulation in this section can be found in Dubois et al. (2014), I give only a brief description in this section.

Horizon-AGN was run for 6 million CPU hours on the Jade SGI Altix supercomputer in France. It simulates a box with periodic boundary conditions, with comoving box width 100 Mpc/h. The simulation has 1024^3 dark matter particles with initial conditions and cosmology from WMAP7 (Komatsu et al., 2011). RAMSES (an adaptive mesh refinement code, Teyssier, 2002) is used to model the gas dynamics, gas cooling and heating, with a minimum cell size of 1 kpc (constant in physical length). Various physical processes are incorporated into the simulation, in particular star formation, feedback from stars (stellar winds, supernovae type II and type Ia), the evolution of six chemical species, and feedback from AGN.

AGN feedback is incorporated into the simulation, and can be either 'radio mode' or 'quasar mode'. Radio mode operates at low black hole accretion rates ($\dot{M}_{BH}/\dot{M}_{Edd} < 0.01$, where $\dot{M}_{BH}$ is the black hole accretion rate, and $\dot{M}_{Edd}$ is the the Eddington accretion rate, see Castelló-Mor et al., 2016), and injects energy into the inter-galactic medium through bipolar jets. Quasar mode operates at higher accretion rates ($\dot{M}_{BH}/\dot{M}_{Edd} > 0.01$), and injects energy isotropically.

Although I do not use it in this thesis, a twin simulation was also run with *no* AGN feedback, the Horizon-no-AGN simulation, making it particularly useful for consistently testing the role of AGN in galaxy formation (Peirani et al., 2016; Beckmann et al., 2017).

6.2.2 Construction of Mock Skies

A light-cone (a simulated box for which one direction is time/redshift) with opening angle of 1 $\deg^2$ was constructed on-the-fly from the simulation with very small redshift steps. Light-cones constructed in this manner mimic the geometry of observational surveys. I use a mock catalogue of galaxies (and associated properties e.g. stellar mass and luminosities) from the light-cone from Laigle et al. (in prep.), constructed in the same way as Kaviraj

et al. (2016). I briefly discuss their process for identifying galaxies within the survey here, but see Kaviraj et al. (2016) for full details.

Galaxies were initially identified from the distribution of star particles using the AdaptaHOP structure finder code (Aubert et al., 2004; Tweed et al., 2009). Selected over-densities must have an over-density of 178 times the average matter density, and at least 50 star particles. Kaviraj et al. (2016) calculated galaxy fluxes using SPS models (see section 1.3.1.4) from Bruzual & Charlot (2003). Each star particle is assumed to behave as an single stellar population, and the total galaxy flux is the sum of the contributions of all star particles. Dust attenuation was calculated using SUNSET[1], using the gas metallicity distribution around the galaxies as a proxy for dust.

The model SPS spectra were then convolved with filters at wavelengths of interest after appropriate redshifting. Laigle et al. constructed a 1 deg^2 mock galaxy catalogue of sources with optical (u, B V r i) and NIR (z, Y, J, H, K) photometry, alongside RA, DEC, stellar mass, star formation rate and redshift. Note that the stellar mass and redshifts used in this chapter are the intrinsic quantities given in the simulation, as opposed to being derived from the mock photometry (but see section 6.4.1 for possible consequences of this choice).

Figure 6.1 shows the masses and redshifts of the galaxies with all $K > 23.5$ galaxies removed as per Chapter 3. We apply the same criteria as Chapter 3 from Johnston et al. (2015) to determine the 90% completeness limit. Apart from the lowest stellar masses in our lowest redshift range (the $M_\star > 10^{9.35} M_\odot$ and $0.5 < z < 0.75$ bin - photometry was only calculated for galaxies with $M_\star > 10^{9.5} M_\odot$), we can consistently compare all our Chapter 3 VIDEO subsamples to the Horizon-AGN mock catalogue.

6.2.3 Measuring Clustering

I measure the clustering, using a HOD model (fitted using MCMC) in Horizon-AGN with almost exactly the same procedure as in Chapter 3. Unlike in Chapter 3, there is no need to remove stars, no mask is needed, and there are no errors on the redshifts, so the weighting system of Arnouts et al. (2002) is unnecessary. The VIDEO field and the mock catalogue are both 1x1 deg^2 fields, so the integral constraint affects both measurements in the same

[1] https://searchcode.com/codesearch/view/91846040/

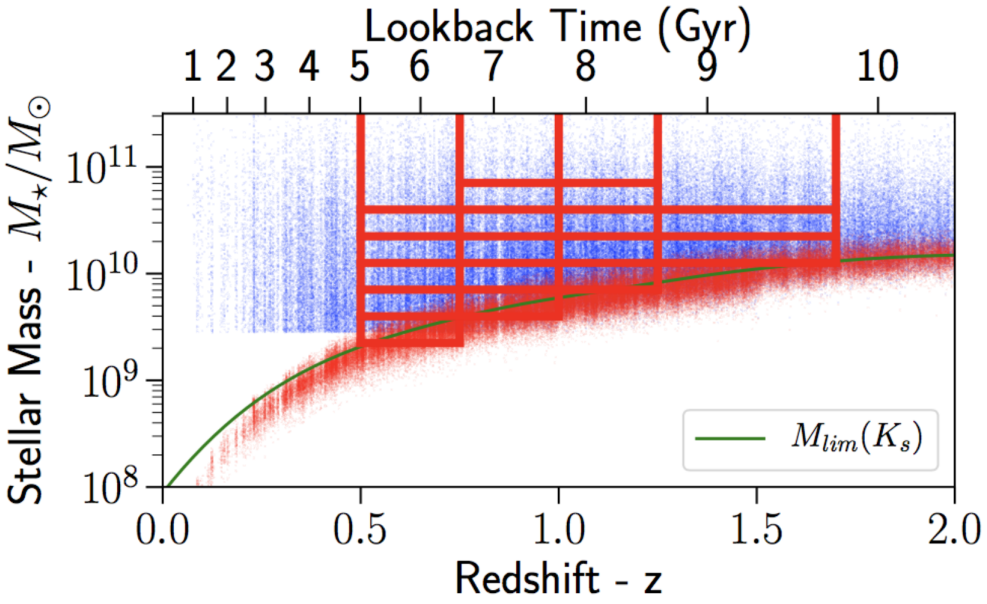


Figure 6.1: Stellar masses and redshifts (blue points) of all $K < 23.5$ galaxies in the Laigle et al. Horizon-AGN mock catalogue. The red lines denote the bins used in Chapter 3. The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of $K_s < 23.5$, and the green curve the implied 90% stellar mass completeness limit, following the approach of Johnston et al. (2015). See figure 4.1 for the corresponding figure in the VIDEO observations.

way.

6.3 Results

I present in this section the clustering measurements, corresponding HOD fits, and comparisons with the equivalent measurements in VIDEO.

6.3.1 ACF Measurements in Horizon-AGN

The clustering measurements from the mock catalogue are shown in Figure 6.2. The essential qualitative behaviour expected based on Chapter 3 is correctly recovered. The mock correlation functions have near power-law behaviour, with amplitudes increasing with stellar mass at all redshifts, and amplitudes of sensible magnitudes. To some degree this is expected - the underlying dark matter distribution is robust between different simulations, and bias is expected to be $\lesssim 2-3$, so it would be extremely surprising if it was orders of mag-

nitude greater or less than expected. But the agreement over linear and non-linear scales should reassure us that Horizon-AGN is correctly qualitatively capturing the large-scale distribution of galaxies.

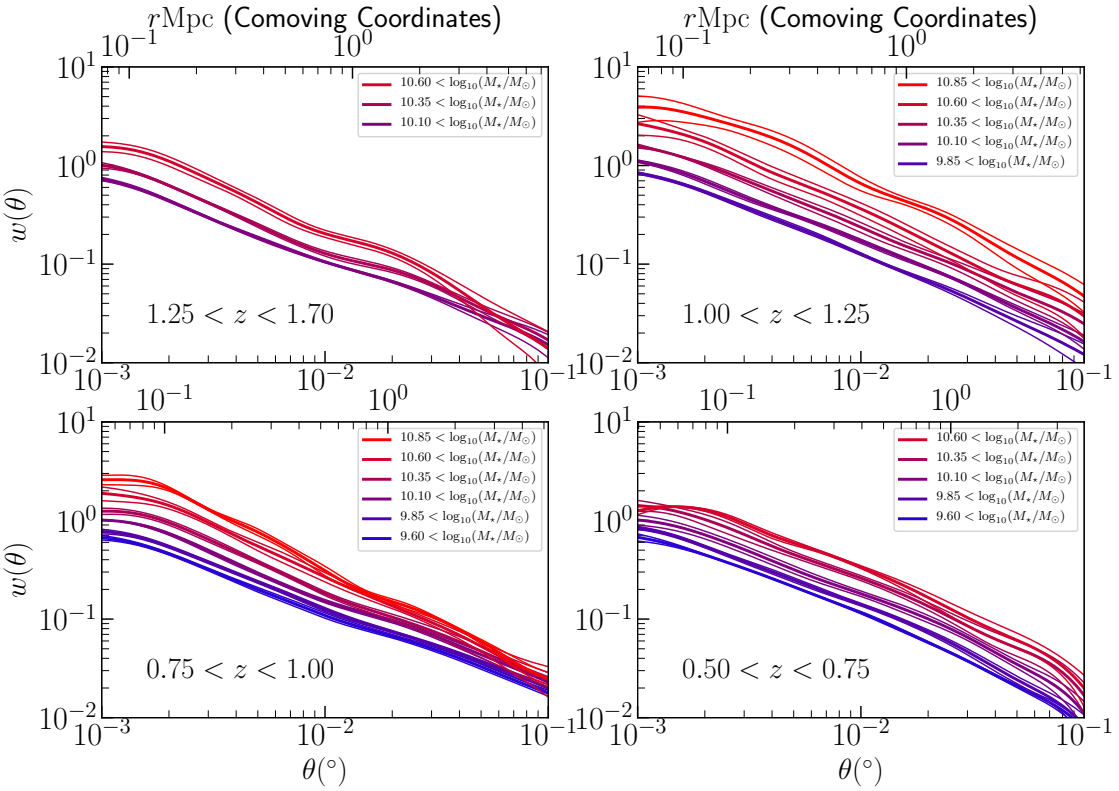


Figure 6.2: The angular correlation function for each stellar mass and redshift bin, measured in the Horizon-AGN mock catalogue.

6.3.2 Comparing VIDEO and Horizon-AGN

I compare my results for each sub-sample of the Horizon-AGN catalogue to the corresponding VIDEO sub-sample measurements from Chapter 3. In particular I plot the ratio of the VIDEO and Horizon-AGN angular correlation functions, see Figure 6.3.

The Horizon-AGN clustering measurements are greater or lesser than the VIDEO measurements by at most a factor of 2 in any bin. In general the trend in each redshift bin is that $\omega_{HORIZON}/\omega_{VIDEO}$ increases with stellar mass e.g. at lower masses Horizon-AGN underestimates the clustering, and at higher stellar masses, overestimates it. The trend over redshift is that at lower redshifts Horizon-AGN underestimates the clustering, and at higher

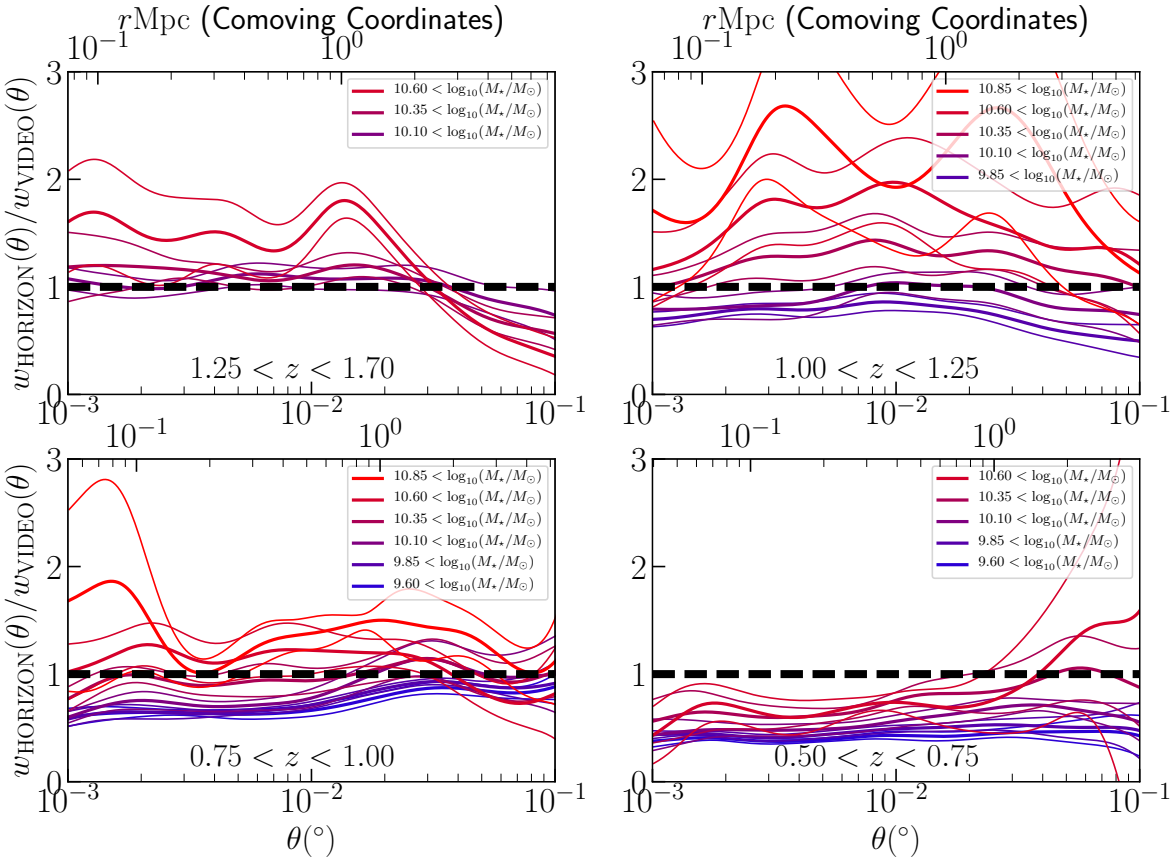


Figure 6.3: Comparison of the VIDEO and Horizon-AGN clustering measurements for the stellar mass and redshift bins. The thin lines are the errors on the ratio, with the errors on both the VIDEO and Horizon-AGN measurements propagated through.The dashed line corresponds to $\omega_{HORIZON} = \omega_{VIDEO}$.

redshifts is closer to the VIDEO clustering, sometimes starting to slightly overestimate it. The underestimation (particularly at low redshifts and stellar masses) has very high statistical significance, however the overestimations (at higher stellar masses and redshifts) typically have lower (1-2σ) significance. In most bins there is very little length scale dependency (the $M_\star > 10^{10.6} M_\odot$, $1.25 < z < 1.7$ bin being the main exception - this appears to be a consequence of that bin in Horizon-AGN having low power at large scales).

6.3.3 HOD Fits

We show in Figure 6.4 the best fit M_{min} and M_1 values from the HOD models fit to the Horizon-AGN measurements, alongside the equivalent VIDEO measurements from Chapter

3. The Horizon-AGN fits give the same qualitative behaviour as observed in VIDEO; M_{min} and M_1 growing as approximate power laws with stellar mass, with M_1 roughly an order of magnitude larger than M_{min}. Horizon-AGN measurements, for both M_{min} and M_1 are generally slightly lower than the equivalent VIDEO measurements. M_{min} values agree well at high stellar masses ($\sim 10^{10.85} M_\odot$), but Horizon-AGN M_{min} values are around 0.5 dex lower than the VIDEO values for stellar masses $\sim 10^{9.6} M_\odot$ - the gradients of the stellar mass to M_{min} lines are slightly different. This is consistent with figure 6.3, with clustering amplitude in Horizon-AGN being much less than that measured in VIDEO at lower stellar masses. M_1 values are consistently a factor of ~ 2 lower for Horizon-AGN than for VIDEO. Similarly to our VIDEO measurements, there appears to be very little evolution in M_{min} and M_1 as a function of stellar mass over the redshift ranges considered here.

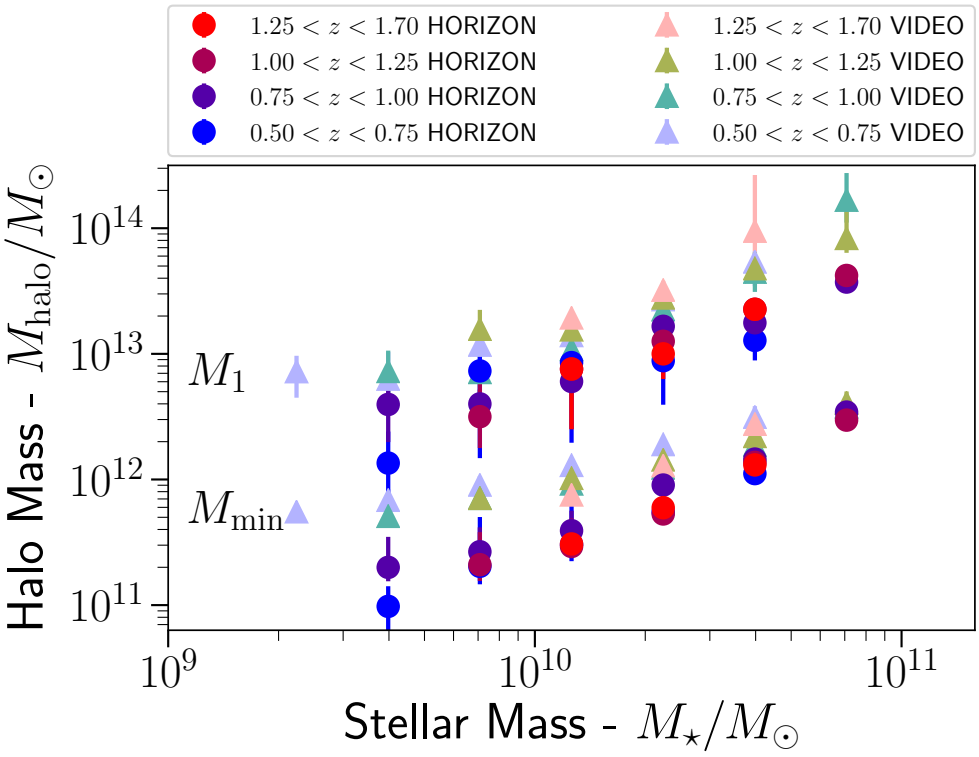


Figure 6.4: Best fit M_{min} and M_1 for the Horizon-AGN clustering measurements, as a function of galaxy sample threshold stellar mass. Equivalent measurements in VIDEO from Chapter 3 shown for comparison.

6.4 Discussion

Causes of discrepancies between observed and simulated measurements can be divided into three categories. Firstly, discrepancies can arise from the simulation inherently not completely capturing the correct galaxy formation physics e.g. the simulation genuinely creates more galaxies of a given mass than are present in the real Universe. Secondly, discrepancies can arise when observational biases are poorly understood and accounted for e.g. the incompleteness of a galaxy survey is underestimated, leading us to falsely conclude that the simulation over produces galaxies, even if the simulation actually matches reality. Finally, there can be differences greater than expected from statistical errors from *cosmic variance* - the galaxy formation physics could be correctly captured in the simulation, but large-scale structure variance could still make measurements from the simulation differ from observations non-trivially.

6.4.1 Observational Biases

A key complexity in comparing results from observations and simulations is that the two 'tools' perceive galaxies in a dramatically different manner. Simulations are concerned with (stellar) mass, whereas observations deal with luminosity. Comparisons between observations and simulations inherently require either inferring mass from observed luminosities, or predicting a luminosity from a simulated mass. Laigle et al., in prep. fit SPS models to the Horizon-AGN mock photometry (essentially treating the photometry as if it was observations) to investigate if the true galaxy masses and redshifts within the simulation can be recovered correctly from the photometry. They find a scatter of ~ 0.2 dex between the 'true' and 'inferred' stellar masses - see Figure 6.5. More problematically, they also find (for the redshifts relevant to this thesis) that the inferred stellar masses are biased to values around ~ 0.2 dex less than the true values. Clearly this is relevant for consistently comparing clustering results between observations and simulations in future. If it is the case that observationally inferred stellar masses are consistently 0.2 dex less than the true values, then comparing the clustering of $M_\star > 10^{10} M_\odot$ observed galaxies with clustering of $M_\star > 10^{10} M_\odot$ simulated galaxies is really comparing the clustering of $M_\star > 10^{10.2} M_\odot$

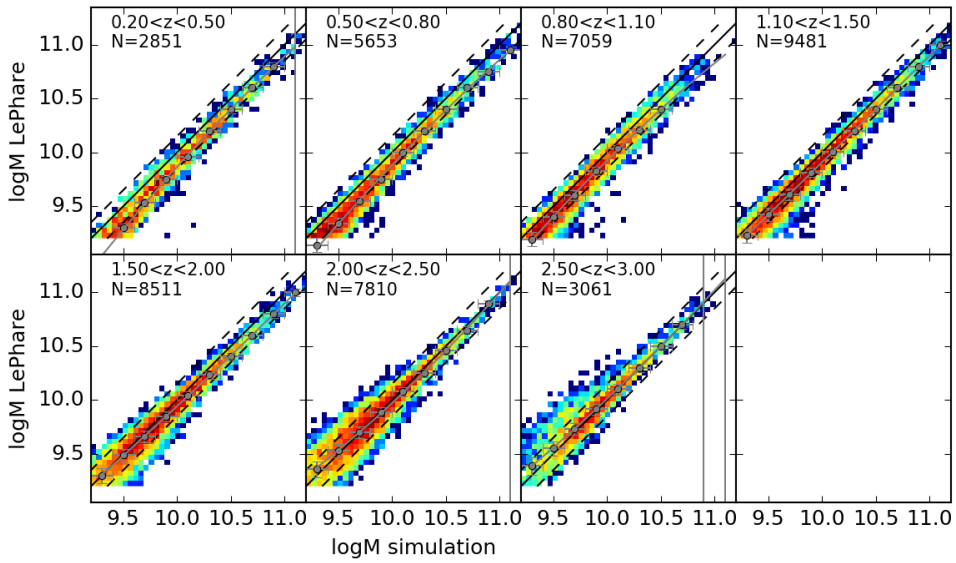


Figure 6.5: Comparison of the galaxy masses direct from the simulation, and as inferred from fitting to the simulated photometry for a range of redshifts. Dashed lines correspond to offsets of 0.2 dex, grey points correspond to the median. Plot from Laigle et al., in prep.

observed galaxies to $M_\star > 10^{10} M_\odot$ simulated galaxies. The higher mass sample will typically have a higher clustering amplitude, so underestimating the stellar masses of observed galaxies will falsely make it appear that the simulation underestimates clustering amplitude (even if it correctly captures galaxy physics).

If stellar mass is consistently underestimated in the observations, then that corresponds to the VIDEO measurements needing to be 'shifted' slightly to the right in Figure 6.4. This would reduce the discrepancy between observations and simulations to a small degree, although the low mass discrepancy described in section 6.4.2 would remain (albeit reduced in size). Future comparisons of clustering between observations and simulations must ensure that stellar masses are being compared consistently.

Beyond the process of inferring stellar mass from photometry, there are several other potential biases in the VIDEO observations that are not a factor in measurements from the Horizon-AGN mock catalogue e.g. errors on photometric redshift and stellar contamination. Future work will focus on better understanding these systematics for more consistent comparisons.

6.4.2 Over Production of Lower Mass Galaxies in Horizon-AGN

Although observational biases may play some role, the main discrepancy in the galaxy-halo connection from clustering between observations and simulations apparent in our results is that the stellar mass to halo mass ratio (SMHR) doesn't drop off as fast in Horizon-AGN as in VIDEO observations.

This raised SMHR for low mass haloes in Horizon-AGN (equivalent to the over production of lower mass galaxies in, or galaxies in lower mass haloes growing too large relative to observations) is reported in Dubois et al. (2014) and Kaviraj et al. (2016). Kaviraj et al. (2016) compare stellar mass functions from the literature to those directly from the simulation, finding an over abundance of galaxies with stellar masses $\lesssim 10^{11} M_{\odot}$. Dubois et al. (2014) compare the SMRH for central galaxies and their haloes directly from the simulation, to observational results from abundance matching in Moster et al. (2013), finding that SMHR observations and simulations agree for haloes with masses $\approx 10^{12} M_{\odot}$, but that galaxies in $\approx 10^{11} M_{\odot}$ haloes are approximately a factor of ten more massive in Horizon-AGN than observations would suggest (which agrees with our results), see Figure 6.6. Kaviraj et al. (2016) suggest that this discrepancy is due to missing physics in the sub-grid supernovae feedback prescription of the simulation. They propose that stronger regulation of star-formation in lower mass haloes could be achieved either through more realistic treatment of the interstellar medium (Kimm et al., 2015), or through stronger winds from star clusters (Agertz & Kravtsov, 2015). Note that the low halo mass regime is one in which using clustering as opposed to merely abundance matching is particularly important for inferring galaxy to halo mass ratios (see Sawala et al., 2015).

Unlike in reality, with simulations it is possible to compare inferences made from mock observations to the true physics within the simulation e.g. it is possible to compare the inferred halo masses from clustering to the true halo masses within the simulation. Figure 6.6 suggests that the inferences from the clustering and HOD modelling agree with the true galaxy-halo relation in the simulation, suggesting that the modelling throughout this thesis is accurate.

Horizon-AGN clustering measurements are also slightly greater than the VIDEO mea-

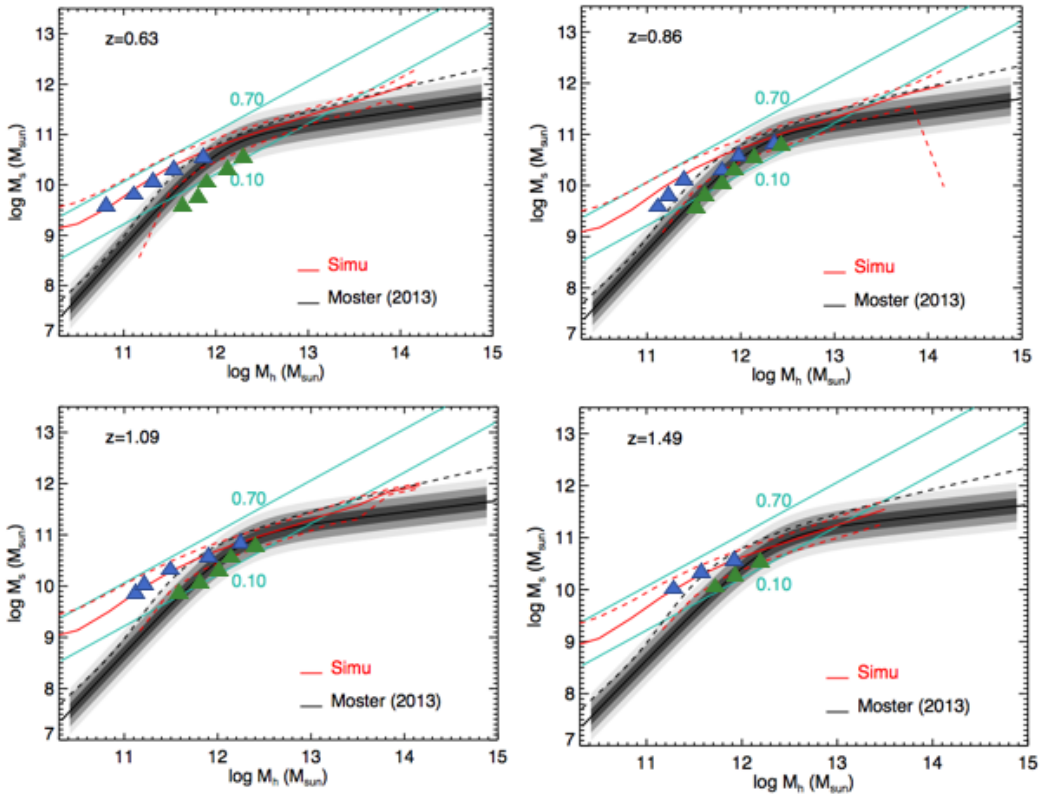


Figure 6.6: Comparison of the central stellar mass to halo mass relation directly from Horizon-AGN (red curve), and from observations (dark shading, from an abundance matching analysis, Moster et al., 2013). Over-plotted are the results from the clustering analysis (green triangles from VIDEO, blue triangles from Horizon-AGN). Teal lines show constant star formation efficiency (of 0.1 and 0.7 respectively). Figure courtesy of Y. Dubois, with data points from this chapter overplotted.

surements for some of the high-redshift and high-stellar mass bins. This corresponds to the slightly lower M_1 values/slightly higher satellite fractions. Future work will investigate if this agrees with the satellite fraction measured directly from the simulation - although the Horizon-AGN measurements are only 1-2σ above the VIDEO results, so it is possible that this small discrepancy is due to cosmic variance.

6.4.3 Cosmic Variance

Cosmic variance (as discussed in Chapters 3 and 5) refers to the observation that statistical measurements can vary by more than Poisson variance due to large-scale structure e.g. the field might sample an extreme over- or under-density of the Universe which can impact on

the statistical properties of the field. For example, as discussed in Chapter 3, there is already known to be moderate difference in the clustering properties of the VIDEO/CFHTLS-D1 field studied in this thesis and the COSMOS field (Durkalec et al., 2015b).

Cosmic variance-like effects, as described in Blaizot et al. (2005), can also impact mock observations derived from simulations in two main ways. Firstly only a finite volume of the Universe is simulated. This requires sampling initial conditions from some larger distribution - a different sampled set of initial conditions would lead to a different simulated universe. Secondly, for a given simulation, constructing the mock sky 'from different viewpoints' will give different sets of mock observations.

As discussed in Chapter 3, VIDEO has three separate fields, so in future work it should be possible to directly estimate the effects of cosmic variance on our observational results. To understand cosmic variance in the simulated results, ideally one would run the simulation multiple times, but this is currently unfeasible for hydrodynamical simulations (although it is more viable in SAMs e.g. Stringer et al., 2009). The approach of sampling different viewpoints within the same simulation (the second type of cosmic variance in simulations discussed above) may be more viable for hydrodynamic simulations, although as Blaizot et al. (2005) discuss, this will generally underestimate the true cosmic variance, as each generated mock sky will still be sampling the same density field, just from different angles.

6.5 Conclusions

In this Chapter I investigated the clustering of galaxies in mock catalogues from the hydrodynamic cosmological simulation Horizon-AGN, to compare to observations, and to confirm that HOD methods give correct physical deductions. I selected mock sources in the simulation that permitted direct comparison to VIDEO sub-samples described in Chapter 3. I then measured the angular correlation function, and fitted HOD models using identical procedures to that applied to the observations, so that a consistent comparison could be made.

I found that the correlation function measurements were in qualitative agreement with the correlation functions observed in VIDEO, producing the correct approximate power law

behaviour, with the correct stellar mass dependance, and the amplitude differing by at most a factor of 2. HOD modelling of the Horizon-AGN measurements also qualitatively recovers the broad nature of the connection between galaxies and haloes suggested by VIDEO, namely M_{min} and M_1 growing as power laws (with a factor of approximately an order of magnitude between them).

The (central) SMHR differs between Horizon-AGN and VIDEO, in agreement with the known result of Horizon-AGN that at low halo masses galaxies are more massive than observations would suggest (equivalently the stellar mass function over produces lower mass galaxies). HOD modelling recovers the galaxy-halo connection as taken directly from the simulation, suggesting that the methodology used throughout this thesis is giving physically correct results.

Future work will focus on investigating if HOD methods correctly recover the satellite population, and the SMHR for galaxies above the knee of the galaxy mass function with the full 12 deg^2 of VIDEO. It will also be possible to compare observations to Horizon-no-AGN to better understand the role of AGN feedback. In particular I plan to use the techniques developed in Chapter 4 to investigate if AGN to star-forming galaxy cross-correlation functions behave differently in Horizon-AGN to Horizon-no-AGN. Future developments will also focus on building a better understanding of the observational biases that need to be incorporated to better compare observations and simulations, including adapting MoMaF (Blaizot et al., 2005) for hydrodynamical simulations.

Chapter 7

Conclusions

In this Chapter I summarise the work in this thesis from Chapters 3-6, and discuss future plans to develop my research.

7.1 Summary

In this thesis I have used multi-wavelength (in particular near-infrared) surveys to trace the large-scale distribution of galaxies in space over cosmic time. Using the 2D location of galaxies on the sky, and the third dimension of redshift as a distance proxy, it is possible to measure the statistical distribution of galaxies over large volumes of the Universe. In particular I focus on quantifying large-scale structure using the angular correlation function, a measure of how much more likely than Poisson two galaxies are to be a given angular separation from each other. I model the angular correlation function using the HOD phenomenology, which breaks clustering down into a) small-scale non-linear clustering of galaxies within individual dark matter haloes, and b) the large-scale linear clustering of the haloes themselves. This process can give an indication of the range of environments the galaxy population is in and the properties of the host halo. I used this model to measure the environment of galaxies with a range of properties at $z \sim 0.5-1.7$ in the VIDEO survey, at $z \sim 6$ in the UDS and UltraVISTA surveys, and compared the measurements to results from the Horizon-AGN hydrodynamical cosmological simulation.

In Chapter 3 I used photometric redshifts and stellar masses from the VIDEO survey

to measure the galaxy-halo relation over $z = 0.5 - 1.7$. I measured the angular correlation function for a range of stellar mass and redshift bins, and fitted HOD models to obtain information about how galaxies occupy halos over cosmic time. My results suggest that there is very little evolution in host halo mass (at fixed stellar mass) over these redshifts. Host halo mass (and the halo mass required to have satellite galaxies) increases with galaxy mass, but does not depend on redshift. I also calculated key derived parameters, in particular finding that bias increases with stellar mass and redshift, reflecting trends in the host halo bias properties. The satellite fraction is reduced for higher redshifts, as there are fewer of the more massive haloes required to host multiple galaxies of a given mass. Finally, I estimated the total stellar mass to halo mass ratio for these redshifts, finding agreement with other studies (in particular UltraVISTA), that the peak is at a halo mass of around $2 \times 10^{12} M_{\odot}$, and is approximately constant over these redshifts

In Chapter 4 I built on my work in Chapter 3 by incorporating physical quenching mechanisms into the HOD model. In particular I introduced a probability of a galaxy being quenched, as a function of halo-centric radius, into the HOD model, which affects how compact the 1-halo term of the clustering is. I fitted this new 'interacting' HOD model to cross-correlation functions between massive galaxies (which will predominantly be centrals), and lower mass galaxies that are either quenched or star-forming, to measure how the lower mass galaxies are quenched in the dense environment associated with the more massive galaxy. In particular I emphasise that on 1-halo scales the cross-correlation function gives information above and beyond the two auto-correlation functions (in particular it measures how the two populations jointly occupy haloes), a property under-appreciated in the literature. I find that at low redshift lower mass galaxies are more likely to be quenched when in the environment of the more massive galaxy, consistent with the literature, but found that at high redshift ($z \sim 1.5$) the quenching of this population has no environmental dependence. Finally I find that the clustering results support 'conformity', that satellites are more likely to be quenched (star-forming) when the central galaxy is quenched (star-forming).

In Chapter 5 I studied the clustering properties of the largest existing sample of extremely bright Lyman-break galaxies at $z \sim 6$. This sample of 263 $M_{UV} \leqslant -20.5$ LBGs

(from Bowler et al., 2015) was selected in the UltraVISTA/COSMOS and UDS/SXDS fields, using deep optical and near-infrared data that is required to distinguish the galaxies from contaminant populations. I measured the clustering of these sources using the angular correlation function, and modelled these measurements within a HOD scheme, incorporating the possibility of the sources having a duty cycle. I found that these bright LBGs ($M_{UV} \leqslant -21$) are extremely highly biased ($b \sim 10$) objects in dense environments, which rules out the possibility of them being fainter ($M_{UV} \approx -19$) galaxies caught in a particularly vigorous period of star formation. This suggests that the bright end of the luminosity function at $z \sim 6$ is determined by feedback processes or dust obscuration, rather than duty cycles. This is of particular interest as Bowler et al. (2014, 2015) report a steepening of the bright end of the ultra-violet luminosity function between $z \sim 6$ and $z \sim 7$. I also find a slight tension between the observed number counts and the measured galaxy bias. This suggests that some slight modification to our understanding of the high-redshift dark matter distribution is needed. I suggest the most likely possibility is that it is necessary to incorporate quasi-linear bias (e.g. Jose et al., 2017).

Finally in Chapter 6 I investigated the ability of the hydrodynamic cosmological simulation Horizon-AGN to recreate the VIDEO clustering measurements from Chapter 3. Using a mock catalogue with the same selection criteria applied to it as to my VIDEO sample, I measured the angular correlation function and fitted HOD models to the simulated data set. I found that Horizon-AGN was largely successful in reproducing clustering measurements, obtaining the correct approximate power law and stellar mass and redshift dependencies (differing in clustering amplitude by at most a factor of 2). The derived stellar mass to halo mass relation from clustering in the simulation exhibited the known low-mass disagreement between Horizon-AGN and observations, namely that galaxies in low mass haloes grow too massive (suggesting that stronger regulation of star-formation is needed in these haloes). In addition, the halo masses from clustering measurements matched those extracted directly from the simulation, giving confidence in the accuracy of my HOD inferences throughout this thesis.

In summary, in this thesis I have shown how clustering measurements of large-scale structure in deep, wide multi-wavelength galaxy surveys can give insight into galaxy evolu-

tion processes over cosmic time. Deep wide-field surveys, such as the VIDEO survey which I focus on in this thesis, allow us to observe statistically significant numbers of galaxies of a range of stellar masses over large volumes of the Universe. Having data from across the electromagnetic spectrum is essential for obtaining stellar masses, photometric redshifts, and estimates of star formation rate. By measuring the angular correlation function, and modelling within a HOD scheme, we can obtain valuable information about the environment galaxies find themselves in, and what physical processes are involved in the build up of their stellar mass. In particular I find that stellar mass to halo mass ratios are constant over the last $\sim$ 7 billion years, and that the role of environment in quenching satellite galaxies appears to increase over this time period. I also used results from cosmological simulations to show that HOD modelling gives accurate information on the galaxy halo connection. I then highlighted how weak regulation of star-formation in low mass haloes in the HORIZON-AGN simulation is reflected in clustering measurements. In the more challenging extremely high redshift regime of $z \sim 6$, I find that the most luminous galaxies are in the densest regions of the Universe, and are not less luminous galaxies caught in a vigorous episode of star formation.

The overarching view of galaxy formation over cosmic time that my results suggest is that of a Universe in which mass quenching begins at very early times, in the first billion years of the Universe, stopping star formation in massive central galaxies, but that environmental quenching only begins to play a role at later times, in the last 6-7 billion years. Future work will focus on developing more sophisticated versions of the HOD model, and applying it to the data sets that will become available in the coming years.

7.2 Future Work

In this section I briefly discuss planned future work that builds on this thesis, in particular the exciting new data sets that will become available in the coming years, and the improved modelling of large-scale structure that I will apply to them.

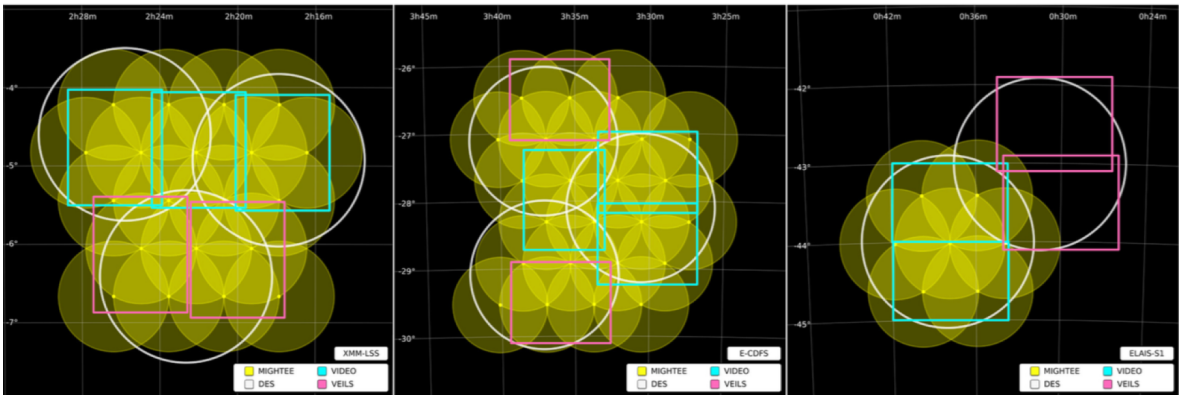


Figure 7.1: Schematic of how the multi-wavelength data over the VIDEO fields will evolve over the coming years. The rectangles are the VISTA paw-prints, teal corresponding to VIDEO, pink to the new VEILS survey. The white circles have deep DES coverage, and the yellow circles are MIGHTEE. Figure from the MIGHTEE proposal (M.Jarvis, private communication)

7.2.1 New Data

There are several new data sets and surveys either currently undergoing observations, or shortly to begin observations, that are relevant to my research.

7.2.1.1 NIR and Optical

Subsequent VIDEO data releases will extend the data to larger areas over three separate fields (eventually 12 deg^2 in total), reducing uncertainty on measurements of the parameters reported in this thesis in future works, extending the angular scales probed by a factor $\sim$3 (allowing for better constraints on both the 1-halo and 2-halo terms), extending to more massive galaxies (allowing better analysis of galaxies above the 'knee' of the stellar mass function), and giving an initial indication of cosmic variance by comparing results between the three fields. This will require the incorporation of deep optical data (in order to obtain high quality photometric redshifts and stellar masses), most likely from the Dark Energy Survey (Abbott et al., 2016) and Hyper Suprime-Cam (Aihara et al., 2017).

In a few years time, the VEILS survey (Hönig et al., 2017) will extend the VIDEO fields in the J and K bands, probing to even larger areas and even higher numbers of sources, see Figure 7.1. By the mid 2020s the *Euclid* space telescope mission will image 15,000 deg^2 of the sky to approximately the same depth as VIDEO - allowing the analyses presented in

this thesis to be extended to an unprecedented level of precision.

7.2.1.2 Radio

The MIGHTEE survey on the MeerKAT telescope (Jarvis, 2012) will give deep radio continuum data over the VIDEO fields (see Figure 7.1). Radio continuum gives two main extra pieces of information about the galaxy population. Firstly, when coupled with the optical and NIR data, the radio continuum can give a less biased estimate of the galaxy star formation rate, see Chapter 1. Secondly, radio data can give an indication of AGN accretion activity. See White et al. (2015) for a discussion of how the two contributions to radio synchrotron emission of AGN activity and star formation can be separated.

To combine radio and optical/NIR data sets, initially I would perform a straightforwards cross match e.g. McAlpine et al. (2012). Then all the analyses described in this thesis could be extended in a fairly straightforwards manner e.g. HOD as a function of AGN radio luminosity. In particular I am interested in applying the techniques developed in Chapter 4 to investigate the role of AGN in quenching e.g. what is the role of AGN in environmental quenching, do AGN heat up gas in a halo, inhibiting star-formation in satellite galaxies for example? To study this I plan to cross-correlate massive galaxies of fixed stellar mass with lower mass star-forming galaxies and compare results when the massive galaxy does/doesn't have an AGN. However a key property of the MIGHTEE survey (and others like it), is that it will have poor resolution relative to its depth and will be confusion limited rather than noise limited; essentially sources will start to overlap long before sources become too faint to be detected. This means that more advanced techniques than straightforwardly cross-matching with an optical survey are needed to make the most of the survey.

Usual approaches to measuring galaxy population statistics like luminosity functions and correlation functions count 5-σ peaks as 'detections', and then calculate statistics on this population. However, this ignores all the information contained within the < 5-σ flux. 'P(D)' Bayesian methods can probe at least two orders of magnitude deeper in luminosity by using the information in the sub-detection flux - essentially a histogram of pixel luminosity is measured, and combined with a noise model, to constrain the source counts or luminosity function below the threshold (e.g. Vernstrom et al., 2014; Zwart et al., 2015). This technique

gives constraints on the 1-point statistics; but Amblard et al. (2011), showed how a related approach can give constraints on the 2-point statistics (the clustering) below the detection limit in Herschel data. By measuring the power spectrum of the unresolved flux, and incorporating a noise model, they were able to constrain the clustering of the unresolved sources, and fit a HOD model to the result, giving the bias and halo masses of sources substantially less luminous than the detection limit.

I have begun work on developing how to implement this in MIGHTEE. I have simulated MIGHTEE continuum images and corresponding power spectra within SKADS (the SKA Design Simulation, Wilman et al., 2008), and am working on modelling these results within a HOD scheme, see figure 7.2. The key result so far is that the unresolved flux power spectrum will only be giving information on extragalactic sources on scales smaller than 0.5 deg - flux on scales larger than that will be dominated by galactic synchrotron. I have also shown that the clustering of radio sources doesn't dramatically impact P(D) analyses for 1-point statistics (which usually assumes the sources aren't clustered) in the regime MIGHTEE will be in by making sources 'overlap' more than might otherwise be expected. This means that measuring the 1- and 2- point statistics do not need to be treated as coupled problems. Future work will focus on implementing this method using real MIGHTEE data.

7.2.2 Improved Modelling

In future analyses of clustering, I will use the HOD parametrisation of Coupon et al. (2015). The model described in Coupon et al. (2015) has the same underlying principles as Zheng et al. (2005) and the approach to modelling described in Chapter 2 (e.g. critical halo mass needed for first galaxy in the halo, number of satellites grows as a power law). However it makes two improvements. Firstly it can model the correlation function of ranges of stellar masses, as opposed to threshold samples. This means that individual galaxies are not included in multiple stellar mass ranges, reducing covariance. Secondly, there is only one model, fitted to all clustering measurements for all stellar masses simultaneously, as opposed to fitting a separate model to each separate stellar mass bin, see figure 7.3.

One advantage of the Coupon et al. (2015) model is that data from different surveys at different angular scales can be used to constrain the same model. With one individual

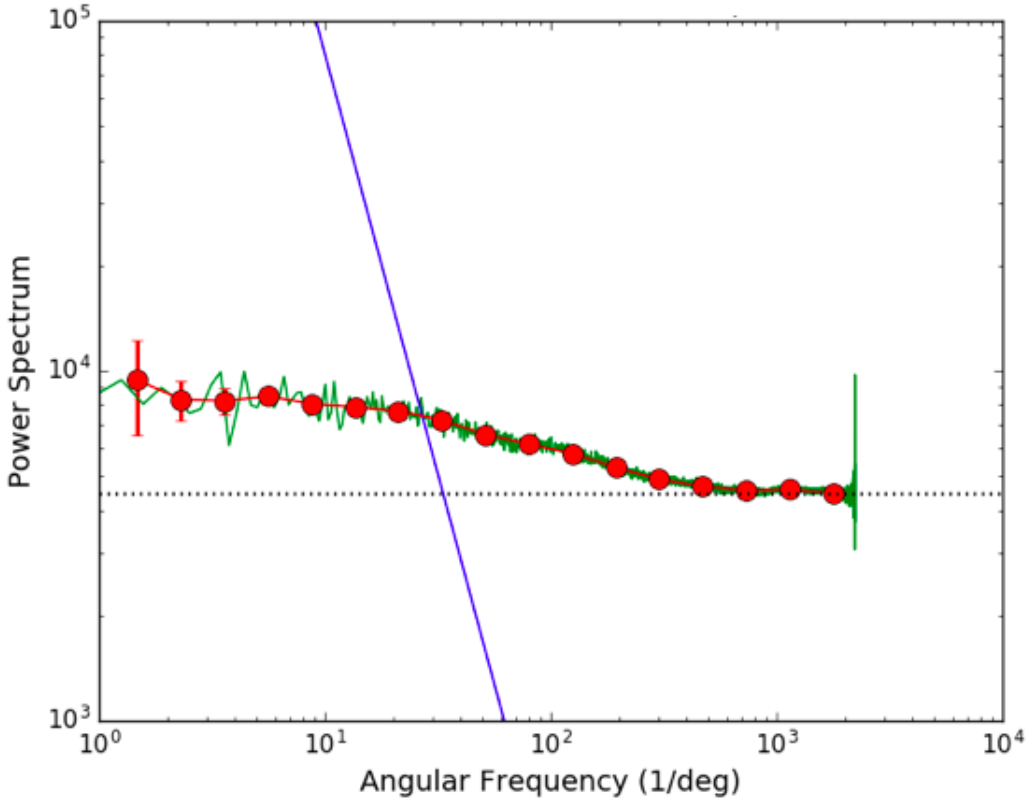


Figure 7.2: Preliminary results of developing methods to measure unresolved flux in MIGHTEE. The x-axis shows inverse angle/angular frequency, and y-axis the power spectrum of the flux. The green line (binned into the red points and error bars) shows the power spectrum of the unresolved flux. The blue line shows the predicted power spectrum from synchrotron radiation from our own Galaxy from Santos et al. (2005). We would anticipate that we would be unable to measure the clustering of the unresolved sources on scales where Galactic synchrotron dominates. The black dotted line shows a fit to the amplitude to the noise - the fact that the power spectrum is not just flat shows that the sources are not uniformly distributed.

survey, one is restricted on the high mass end by volume (galaxies above a given stellar mass are too rare), and on the lower mass end by sensitivity (galaxies below a given stellar mass are too faint). This is the reason for the common 'wedding tier' observing strategy - permitting the full range of stellar masses to be probed. Similarly wide surveys can probe clustering to larger areas, but are typically shallower so cannot probe small scales, and vice versa. Alongside the two VISTA surveys, VIDEO and UltraVISTA, discussed in this thesis (which essentially form the deepest two tiers of a wedding cake), there is a third wider extragalactic tier the VISTA Kilo-degree Infrared Galaxy survey (VIKING, Arnaboldi et al.,

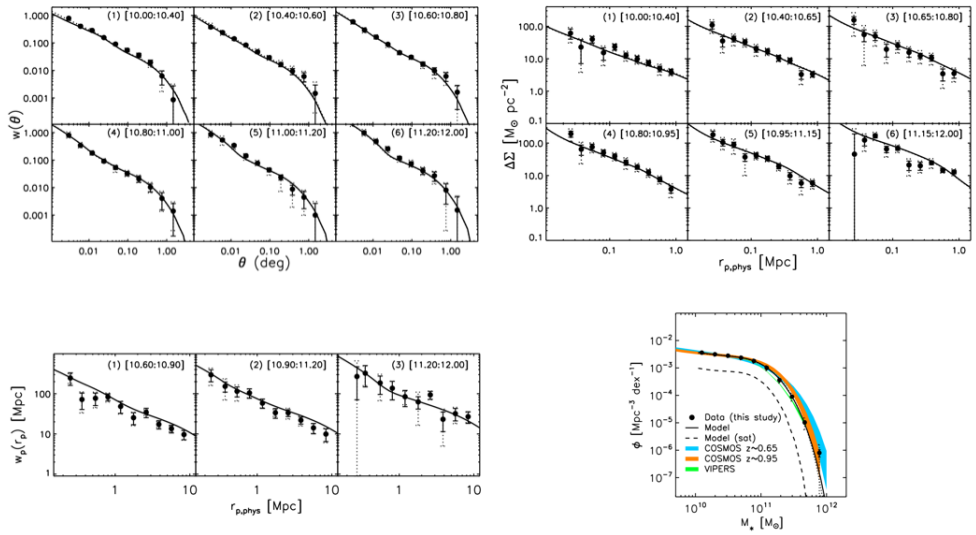


Figure 7.3: All the measurements in Coupon et al. (2015) - stellar mass function, clustering (angular and projected) for several stellar mass ranges, and galaxy-galaxy lensing. The black curve that goes through all these points is all from just one model simultaneously fitted to all the data.

2007). Figure 7.4 shows schematically what clustering measurements in these three tiers would be expected to look like. Modelling clustering in all three surveys at once would allow a model to span many more decades of stellar mass and angular scale than would be possible just modelling data in one survey. Similarly, this approach would allow the clustering of the lower luminosity LBGs of Harikane et al. (2016) and the brighter LBGs of Bowler et al. (2015) used in Chapter 5 to be simultaneously modelled.

7.2.2.1 Incorporating More Physical Processes into HOD Models

Ideally we would want to model both occupation numbers and galaxy quenching processes (see Chapter 4) simultaneously in one model. In the case of developing Scenario 2 type models as described in Chapter 4, I will parametrise the probability of being quenched as a function of halo mass and radius from centre of halo, then fit these parameters alongside the Coupon et al. (2015) parameters, simultaneously to i) passive auto-correlation functions stellar mass ranges, ii) star-forming auto-correlation functions and iii) relevant cross-correlations.

To investigate conformity I would then introduce two different parameterisations for the

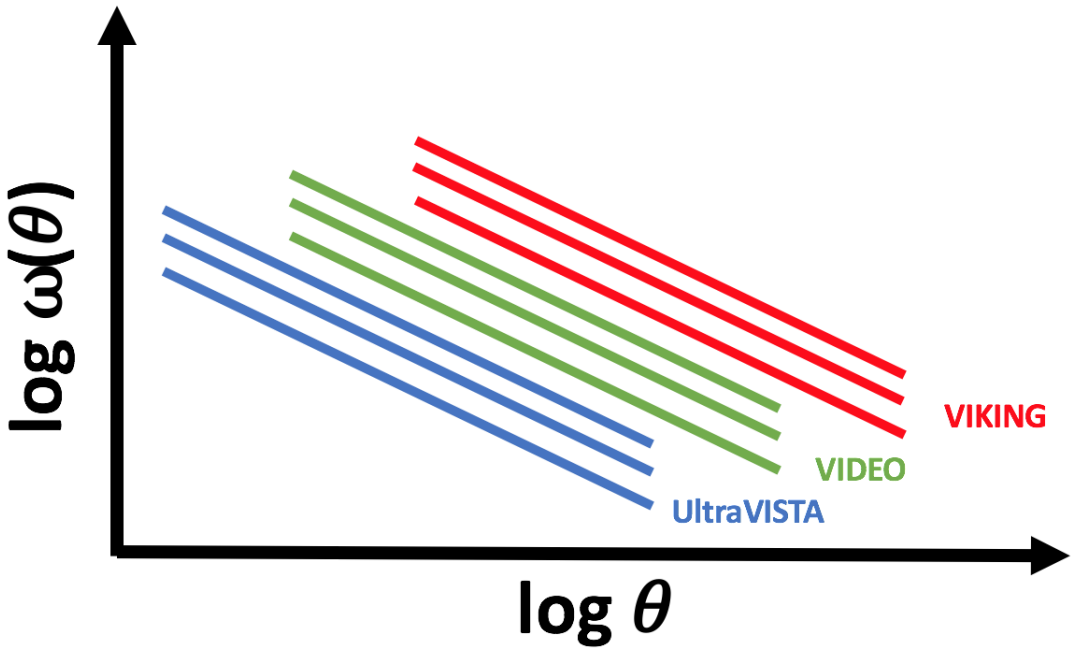


Figure 7.4: Cartoon of combining data from three wide deep extra-galactic surveys at one redshift (assuming that clustering continues to increase with stellar mass over the relevant mass scales). Using the Coupon et al. (2015) model, all of this data (at one redshift) could go into constraining one model, rather than a separate model at each stellar mass.

probability of satellites being quenched, one for if the central is quenched, one for if the central is star forming. Beyond that, further parameters could be included to account for if the central galaxy hosts an AGN etc.

With regards to our understanding of the early Universe (and of how to incorporate reionization into measurements of LSS), VEILS and *Euclid*, will allow extension to even more luminous LBGs at high redshift, giving improved constraints on the luminosity function and clustering of high-redshift galaxies. By the mid 2020s it should be possible to use *Euclid* to do similar analyses to those presented in this thesis with 1,000s of LBGs (Bowler et al., 2017), which will reveal how the measured large scale structure of LBGs and LAEs relates to reionization.

7.2.3 Non-Catalogue Information

Throughout this thesis I fit models using information from two types of catalogue information - clustering and number counts. However Coupon et al. (2015), as discussed earlier, also

fit using information from galaxy-galaxy lensing shear measurements, see figure 7.3. Over the VIDEO fields there are CHFTLens shear measurements (see Heymans et al., 2012). In future I plan to fit HOD models to both clustering measurements and galaxy-galaxy lensing from CHFTLens over the XMM-LSS VIDEO fields.

There are however alternative astronomical measurements that can give an indication of the galaxy host halo environment. Measurements of the X-ray luminosity and temperature of hot gas in haloes can give estimates of halo mass (e.g. Anderson et al., 2015). The Sunayev-Zeldovich effect, as discussed in Chapter 1, gives an indication of halo mass and is particularly effective at high redshift as the phenomenon has little redshift dependence (see Ade et al., 2013). Even HI rotation curves, Díaz-García et al. (2016), may be able to contribute to constraining the galaxy-halo connection (HI is needed as the rotation curves need to probe to large radii to measure the mass of the whole halo). MIGHTEE will be able to measure rotation curves for the most massive HI galaxies up to $z \sim 0.5$.

In general I plan to extend Coupon et al. (2015) by adding extra likelihood terms to constrain the galaxy-halo connection, beyonds the likelihoods from clustering and lensing, e.g.:

$$\chi^2 = \chi^2_\phi + \Sigma\chi^2_{\omega(\theta)} + \Sigma\chi^2_{\Delta\Sigma(r)} + \Sigma\chi^2_X \tag{7.1}$$

where χ^2_ϕ is the likelihood from number counts, $\Sigma\chi^2_{\omega(\theta)}$ is the sum of likelihoods from all clustering measurements, $\Sigma\chi^2_{\Delta\Sigma(r)}$ is the sum of likelihoods from all lensing measurements, and $\Sigma\chi^2_X$ is the likelihood from constraints on halo masses from X-ray luminosity measurements, or any other probe. This approach of parametrising all probes within the same model also makes checking for consistency easier e.g. would permit one to tell if X-ray estimates of halo masses are consistently higher than those from clustering, or the satellite fraction inferred from galaxy-galaxy lensing is consistently lower than from clustering etc. Using the extra information contained in the alternative probes should improve constraints and break degeneracies in the modelling.

7.2.4 Going from a Static to a Dynamic View of Halo Occupation

The view presented of the galaxy-halo connection by halo occupation modelling and photometric surveys as outlined in this thesis gives a 'static' image of how galaxies occupy dark matter halos over cosmic time. However, it is also extremely important to understand the dynamics of the occupancy as it can be linked to how recently mergers have occurred, and how fast virialization occurs in different halos - crucial at this epochs we are interested in for this thesis ($z \sim 1-2$, the peak of star formation). An important way of probing these dynamics is to measure the redshift space distortions (as described in section 2.1.1.1) of the correlation function in a spectroscopic survey, which gives a measure of velocity dispersion inside a halo (Guo et al., 2016).

Key upcoming spectroscopic surveys and instruments relevant for this science goal include:

- The Multi-Object Optical and Near-infrared Spectrograph (MOONS, Cirasuolo et al., 2014), a VLT instrument due to start observations in ∼2019. MOONS currently has plans to survey all VIDEO sources at $z > 1$ with $H < 23.5$, which should allow exquisite measurement of redshift-space distortions for optical/NIR selected galaxies on both sides of the knee of the stellar mass function.

- WEAVE-LOFAR (Smith et al., 2016) is a survey starting observations in 2018 following up Low-Frequency Array (LOFAR, Röttgering et al., 2011; Williams et al., 2016) sources with the optical spectrograph WEAVE (Dalton et al., 2012). It will be possible to measure the redshift space distortion of radio selected sources at high redshift for the first time, giving an indication of the velocity dispersion of star forming galaxies within halos, and measure how stationary central AGN are.

- Slightly further in the future the Prime Focus Spectrograph (PFS, on the Subaru telescope, Wang et al., 2016) and the 4m Multi Object Spectroscopic Telescope (4MOST, on VISTA, Depagne, 2015) instruments will begin observations in 2019 and 2021 respectively. Both are likely to cover a substantial subset of the VIDEO fields, and will give a magnificent view of the dynamics of the galaxy-halo connection at $z = 1-2$.

Bibliography

Abadi M. G., Moore B., Bower R. G., 1999, Monthly Notices of the Royal Astronomical Society, 308, 947

Abbas U. et al., 2010, Monthly Notices of the Royal Astronomical Society, 406, 1306

Abbott T. et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1270

Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, Monthly Notices of the Royal Astronomical Society, 417, 1891

Adam R. et al., 2016, Astronomy & Astrophysics, 596, A108

Ade P. A. R. et al., 2013, Astronomy & Astrophysics, 557, A52

Ade P. A. R. et al., 2016, Astronomy & Astrophysics, 594, A13

Agertz O., Kravtsov A. V., 2015, The Astrophysical Journal, 804, 18

Aihara H. et al., 2017, eprint arXiv:1702.08449

Allison R. et al., 2015, Monthly Notices of the Royal Astronomical Society, 451, 849

Alpher R. A., Bethe H., Gamow G., 1948, Physical Review, 73, 803

Amblard A. et al., 2011, Nature, 470, 510

Anderson L. et al., 2014, Monthly Notices of the Royal Astronomical Society, 441, 24

Anderson M. E., Gaspari M., White S. D. M., Wang W., Dai X., 2015, Monthly Notices of the Royal Astronomical Society, 449, 3806

Antolini C., Martinelli M., Fantaye Y., Baccigalupi C., 2013, Journal of Cosmology and Astroparticle Physics, 2013, 024

Antonuccio-Delogu V., Silk J., 2010, Monthly Notices of the Royal Astronomical Society, 405, 1303

Arnaboldi M., Neeser M. J., Parker L. C., Rosati P., Lombardi M., Dietrich J. P., Hummel W., 2007, The Messenger, Vol. 127. European Southern Observatory

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, Monthly Notices of the Royal Astronomical Society, 310, 540

Arnouts S. et al., 2002, Monthly Notices of the Royal Astronomical Society, 329, 355

Artale M. C. et al., 2016, eprint arXiv:1611.05064

Asorey J., Carrasco Kind M., Sevilla-Noarbe I., Brunner R. J., Thaler J., 2016, Monthly Notices of the Royal Astronomical Society, 459, 1293

Aubert D., Pichon C., Colombi S., 2004, Monthly Notices of the Royal Astronomical Society, 352, 376

Baade W., Zwicky F., 1938, The Astrophysical Journal, 88, 411

Bacon D. J., Refregier A. R., Ellis R. S., 2000, Monthly Notices of the Royal Astronomical Society, 318, 625

Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, Monthly Notices of the Royal Astronomical Society, 373, 469

Baldry I. K. et al., 2010, Monthly Notices of the Royal Astronomical Society, 404, 86

Ball N. M., Loveday J., Brunner R. J., 2007, Monthly Notices of the Royal Astronomical Society, 383, 907

Balogh M. et al., 2004, Monthly Notices of the Royal Astronomical Society, 348, 1355

Banerji M. et al., 2014, Monthly Notices of the Royal Astronomical Society, 446, 2523

Barcons X., 2001, in AIP Conference Proceedings, Vol. 599, AIP, pp. 3–12

Barone-Nugent R. L. et al., 2014, The Astrophysical Journal, 793, 17

Bassett B. A., Hlozek R., 2010, Dark Energy: Observational and Theoretical Approaches, ISBN 9780521518888, 246

Bastian N., Covey K. R., Meyer M. R., 2010, Annual Review of Astronomy and Astrophysics, 48, 339

Becker G. D., Bolton J. S., Lidz A., 2015, Publications of the Astronomical Society of Australia, 32, e045

Beckmann R. S. et al., 2017, eprint arXiv:1701.07838

Behroozi P., Silk J., 2016, eprint arXiv:1609.04402

Behroozi P. S., Wechsler R. H., Conroy C., 2013, The Astrophysical Journal, 770, 57

Bell E. F. et al., 2004, The Astrophysical Journal, 608, 752

Benot A., 2002, in AIP Conference Proceedings, Vol. 616, AIP, pp. 31–38

Benson A. J., 2010, Physics Reports, 495, 58

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 1999, Monthly Notices of the Royal Astronomical Society, Volume 311, Issue 4, pp. 793-808., 311, 793

Bergvall N., Marquart T., Way M. J., Blomqvist A., Holst E., Östlin G., Zackrisson E., 2016, Astronomy & Astrophysics, 587, A72

Berlind A. A., Weinberg D. H., 2002, The Astrophysical Journal, 575, 587

Berlind A. A. et al., 2003, The Astrophysical Journal, 593, 1

Berti A. M., Coil A. L., Behroozi P. S., Eisenstein D. J., Bray A. D., Cool R. J., Moustakas J., 2017, The Astrophysical Journal, 834, 87

Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393

Betoule M. et al., 2014, Astronomy & Astrophysics, 568, A22

Beutler F. et al., 2011, Monthly Notices of the Royal Astronomical Society, 416, 3017

Bielby R. M. et al., 2014, Astronomy & Astrophysics, 568, A24

Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, Monthly Notices of the Royal Astronomical Society, 360, 159

Blanton M., Cen R., Ostriker J. P., Strauss M. A., 1999, The Astrophysical Journal, 522, 590

Blanton M., Cen R., Ostriker J. P., Strauss M. A., Tegmark M., 2000, The Astrophysical Journal, 531, 1

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Blumenthal G. R., Pagels H., Primack J. R., 1982, Nature, 299, 37

Bond J. R., Efstathiou G., 1984, The Astrophysical Journal, 285, L45

Boulade O. et al., 2003, in Instrument Design and Performance for Optical/Infrared Groundbased Telescopes, Proceedings of the SPIE, Iye M., Moorwood A. F. M., eds., Vol. 4841, p. 72

Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, The Astrophysical Journal, 670, 928

Bouwens R. J. et al., 2015, The Astrophysical Journal, 803, 34

Bouwens R. J. et al., 2016, The Astrophysical Journal, 830, 67

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, Monthly Notices of the Royal Astronomical Society, 370, 645

Bowler R. A. A. et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 1817

Bowler R. A. A., Dunlop J. S., McLure R. J., McLeod D. J., 2017, Monthly Notices of the Royal Astronomical Society, 466, 3612

Bowler R. A. A. et al., 2014, Monthly Notices of the Royal Astronomical Society, 440, 2810

Bowman A. W., 1984, Biometrika, 71, 353

Branch D., 1998, Annual Review of Astronomy and Astrophysics, 36, 17

Bruzual G., Charlot S., 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000

Bull P. et al., 2016, Physics of the Dark Universe, 12, 56

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, Monthly Notices of the Royal Astronomical Society, 321, 559

Burgarella D. et al., 2013, Astronomy & Astrophysics, 554, A70

Campanelli L., Fogli G. L., Kahniashvili T., Marrone A., Ratra B., 2012, The European Physical Journal C, 72, 2218

Carroll S. M., Press W. H., Turner E. L., 1992, Annual Review of Astronomy and Astrophysics, 30, 499

Cassata P. et al., 2007, The Astrophysical Journal Supplement Series, 172, 270

Castelló-Mor N., Netzer H., Kaspi S., 2016, Monthly Notices of the Royal Astronomical Society, 458, 1839

Cen R., Ostriker J. P., 2000, The Astrophysical Journal, 538, 83

Cen R., Safarzadeh M., 2014, The Astrophysical Journal, 798, L38

Chisari N. et al., 2015, Monthly Notices of the Royal Astronomical Society, 454, 2736

Chiu I. et al., 2016, Monthly Notices of the Royal Astronomical Society, 458, 379

Choi J., Conroy C., Moustakas J., Graves G. J., Holden B. P., Brodwin M., Brown M. J. I., van Dokkum P. G., 2014, The Astrophysical Journal, 792, 95

Ciambur B. C., Kauffmann G., Wuyts S., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2488

Cirasuolo M. et al., 2014, in Ground-based and Airborne Instrumentation for Astronomy V, Proceedings of the SPIE, Ramsay S. K., McLean I. S., Takami H., eds., International Society for Optics and Photonics, p. 91470N

Clay S. J., Thomas P. A., Wilkins S. M., Henriques B. M. B., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2692

Clowe D., Gonzalez A., Markevitch M., 2004, The Astrophysical Journal, 604, 596

Clowes R. G., Harris K. A., Raghunathan S., Campusano L. E., Sochting I. K., Graham M. J., 2013, Monthly Notices of the Royal Astronomical Society, 429, 2910

Coe D., 2010, eprint arXiv:1005.0411

Cole S., Kaiser N., 1989, Monthly Notices of the Royal Astronomical Society, 237, 1127

Condon J. J., 1992, Annual Review of Astronomy and Astrophysics, 30, 575

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, The Astrophysical Journal, 647, 201

Cooray A., Sheth R., 2002, Physics Reports, 372, 1

Coupon J. et al., 2015, Monthly Notices of the Royal Astronomical Society, 449, 1352

Coupon J. et al., 2012, Astronomy & Astrophysics, 542, A5

Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L., 1996, The Astrophysical Journal, 473, 7

Croft R. A. C., 2013, Monthly Notices of the Royal Astronomical Society, 434, 3008

Croton D., Darren, 2007, Bulletin of the American Astronomical Society, 39, 955

Croton D. J. et al., 2006, Monthly Notices of the Royal Astronomical Society, 365, 11

Croton D. J. et al., 2016, The Astrophysical Journal Supplement Series, 222, 22

Dalton G. et al., 2012, in Proceedings of the SPIE, McLean I. S., Ramsay S. K., Takami H., eds., Vol. 8446, p. 12

Dalton G. B. et al., 2006, in Ground-based and Airborne Instrumentation for Astronomy, Proceedings of the SPIE, McLean I. S., Iye M., eds., Vol. 6269, p. 62690X

Darvish B., Mobasher B., Sobral D., Rettura A., Scoville N., Faisst A., Capak P., 2016, eprint arXiv:1605.03182

Darvish B., Mobasher B., Sobral D., Scoville N., Aragon-Calvo M., 2015, The Astrophysical Journal, 805, 121

Davé R., Thompson R., Hopkins P. F., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3265

Davidzon I. et al., 2017, eprint arXiv:1701.02734

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, The Astrophysical Journal, 292, 371

Davis M., Geller M. J., 1976, The Astrophysical Journal, 208, 13

Dawson K. S. et al., 2016, The Astronomical Journal, 151, 44

de Blok W. J. G., 2010, Advances in Astronomy, 2010, 1

Dekel A., Silk J., 1986, The Astrophysical Journal, 303, 39

Del Popolo A., Pace F., 2016, Astrophysics and Space Science, 361, 162

Depagne É., 2015, in Astrophysics and Space Science Proceedings, Vol. 39, Springer International Publishing, pp. 147–154

Desert F. X., 2000, eprint arXiv:astro-ph/0006360

Díaz-García S., Salo H., Laurikainen E., Leaman R., 2016, eprint arXiv:1611.01844

Doré O., Holder G., Alvarez M., Iliev I. T., Mellema G., Pen U.-L., Shapiro P. R., 2007, Physical Review D, 76, 043002

Draine B., 2003, Annual Review of Astronomy and Astrophysics, 41, 241

Dressler A., 1980, The Astrophysical Journal, 236, 351

Driver S. P. et al., 2011, Monthly Notices of the Royal Astronomical Society, 413, 971

Drory N. et al., 2009, The Astrophysical Journal, 707, 1595

Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, Monthly Notices of the Royal Astronomical Society, 463, 3948

Dubois Y. et al., 2014, Monthly Notices of the Royal Astronomical Society, 444, 1453

Duivenvoorden S. et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 277

Dunlop J. S. et al., 2013, Monthly Notices of the Royal Astronomical Society, 432, 3520

Dupé F.-X., Rassat A., Starck J.-L., Fadili M. J., 2011, Astronomy & Astrophysics, 534, A51

Durkalec A. et al., 2015a, Astronomy & Astrophysics, 576, L7

Durkalec A. et al., 2015b, Astronomy & Astrophysics, 583, A128

Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, Monthly Notices of the Royal Astronomical Society, 235, 715

Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705

Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 844

Eisenstein D. J., Seo H.-j., White M., 2007, The Astrophysical Journal, 664, 660

Elbaz D. et al., 2007, Astronomy and Astrophysics, 468, 33

Ellis R. S., 2010, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368

Fabian A., 2012, Annual Review of Astronomy and Astrophysics, 50, 455

Faessler A., Hodák R., Kovalenko S., Šimkovic F., 2017, International Journal of Modern Physics E, 26, 1740008

Farouki R., Shapiro S. L., 1981, The Astrophysical Journal, 243, 32

Feain I. J., Papadopoulos P. P., Ekers R. D., Middelberg E., 2007, The Astrophysical Journal, 662, 872

Fierlinger K. M., Burkert A., Ntormousi E., Fierlinger P., Schartmann M., Ballone A., Krause M. G. H., Diehl R., 2016, Monthly Notices of the Royal Astronomical Society, 456, 710

Fine S., Shanks T., Johnston R., Jarvis M. J., Mauch T., 2015, Monthly Notices of the Royal Astronomical Society, 452, 2692

Finkelstein S. L. et al., 2015, The Astrophysical Journal, 810, 71

Firth R. E. et al., 2014, Monthly Notices of the Royal Astronomical Society, 446, 3895

Fischera J., Dopita M. A., Sutherland R. S., 2003, The Astrophysical Journal, 599, L21

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of the Pacific, 125, 306

Frenk C., White S., 2012, Annalen der Physik, 524, 507

Friedmann A., 1922, Zeitschrift für Physik, 10, 377

Fukugita M., Hogan C. J., Peebles P. J. E., 1998, The Astrophysical Journal, 503, 518

Fukugita M., Peebles P. J. E., 2004, The Astrophysical Journal, 616, 643

Furlong M. et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 4486

Furusawa H. et al., 2008, The Astrophysical Journal Supplement Series, 176, 1

Gabor J. M., Dave R., 2014, Monthly Notices of the Royal Astronomical Society, 447, 374

Geach J. E. et al., 2013, The Astrophysical Journal, 776, L41

Giavalisco M., 2002, Annual Review of Astronomy and Astrophysics, 40, 579

Grissom R. L., Ballantyne D. R., Wise J. H., 2014, Astronomy & Astrophysics, 561, A90

Groth E. J., Peebles P. J. E., 1977, The Astrophysical Journal, 217, 385

Guhathakurta P., Tyson J. A., Majewski S. R., 1990, The Astrophysical Journal, 357, L9

Gunn J. E., Gott J. R. I., 1972, The Astrophysical Journal, 176, 1

Guo H., Zehavi I., Zheng Z., 2012, The Astrophysical Journal, 756, 127

Guo H. et al., 2016, Monthly Notices of the Royal Astronomical Society, 459, 3040

Gwyn S. D. J., 2012, The Astronomical Journal, 143, 38

Han J., Cole S., Frenk C. S., Jing Y., 2016, Monthly Notices of the Royal Astronomical Society, 457, 1208

Harikane Y. et al., 2016, The Astrophysical Journal, 821, 123

Hartley W. G., Almaini O., Foucaud S., 2013, in Thirty Years of Astronomical Discovery with UKIRT, Astrophysics and Space Science Proceedings, Vol. 37, Springer Science+Business Media, pp. 309–321

Hartley W. G., Almaini O., Mortlock A., Conselice C., 2012, Proceedings of the International Astronomical Union, 8, 105

Hartley W. G., Conselice C. J., Mortlock A., Foucaud S., Simpson C., 2015, Monthly Notices of the Royal Astronomical Society, 451, 1613

Hastings W. K., 1970, Biometrika, 57, 97

Hatfield P. W., Bowler R. A. A., Jarvis M. J., Hale C. L., 2017, eprint arXiv:1702.03309

Hatfield P. W., Jarvis M. J., 2016, eprint arXiv:1606.08989

Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, Monthly Notices of the Royal Astronomical Society, 459, 2618

Hayward C. C., Smith D. J. B., 2014, Monthly Notices of the Royal Astronomical Society, 446, 1512

Hearin A. P., Behroozi P. S., van den Bosch F. C., 2016a, Monthly Notices of the Royal Astronomical Society, 461, 2135

Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016b, Monthly Notices of the Royal Astronomical Society, 460, 2552

Heymans C. et al., 2012, Monthly Notices of the Royal Astronomical Society, 427, 146

Hillebrandt W., Niemeyer J. C., 2000, Annual Review of Astronomy and Astrophysics, 38, 191

Hinshaw G. et al., 2013, The Astrophysical Journal Supplement Series, 208, 19

Hirschmann M., De Lucia G., Wilman D., Weinmann S., Iovino A., Cucciati O., Zibetti S., Villalobos A., 2014, Monthly Notices of the Royal Astronomical Society, 444, 2938

Hogg D. W., 1999, eprint arXiv:9905116

Hogg D. W. et al., 2003, The Astrophysical Journal, 585, L5

Hönig S. F. et al., 2017, Monthly Notices of the Royal Astronomical Society, 464, 1693

Hubble E., 1929, Proceedings of the National Academy of Sciences, 15, 168

Hubble E. P., 1926, The Astrophysical Journal, 64, 321

Hudson M. J., Harris G. L., Harris W. E., 2014, The Astrophysical Journal, 787, L5

Ilbert O. et al., 2015, Astronomy & Astrophysics, 579, A2

Ilbert O. et al., 2006, Astronomy and Astrophysics, 457, 841

Jackson J. C., 1972, Monthly Notices of the Royal Astronomical Society, 156, 1P

Jarvis M. J., 2012, African Skies, Vol. 16, p.44, 16, 44

Jarvis M. J. et al., 2013, Monthly Notices of the Royal Astronomical Society, 428, 1281

Johnston R., 2011, The Astronomy and Astrophysics Review, 19, 41

Johnston R., Vaccari M., Jarvis M., Smith M., Giovannoli E., Häußler B., Prescott M., 2015, Monthly Notices of the Royal Astronomical Society, 453, 2541

Jose C., Baugh C. M., Lacey C. G., Subramanian K., 2017, eprint arXiv:1702.00853

Jose C., Lacey C. G., Baugh C. M., 2016, Monthly Notices of the Royal Astronomical Society, 463, 270

Jose C., Srianand R., Subramanian K., 2013, Monthly Notices of the Royal Astronomical Society, 435, 368

Jullo E., Kneib J.-P., Limousin M., Elíasdóttir Á., Marshall P. J., Verdugo T., 2007, New Journal of Physics, 9, 447

Kaiser N., 1987, Monthly Notices of the Royal Astronomical Society, 227, 1

Kamada A., Inoue K. T., Takahashi T., 2016, Physical Review D, 94, 023522

Kappl R., Winkler M. W., 2011, Nuclear Physics B, 850, 505

Katsianis A., Tescari E., Wyithe J. S. B., 2016, Publications of the Astronomical Society of Australia, 33, e029

Kauffmann G., Li C., Zhang W., Weinmann S., 2013, Monthly Notices of the Royal Astronomical Society, 430, 1447

Kauffmann G., Nusser A., Steinmetz M., 1997, Monthly Notices of the Royal Astronomical Society, 286, 795

Kaviraj S. et al., 2016, Monthly Notices of the Royal Astronomical Society, stx126

Kawinwanichakij L. et al., 2016, The Astrophysical Journal, 817, 9

Kerscher M., Szapudi I., Szalay A. S., 2000, The Astrophysical Journal, 535, L13

Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1349

Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901

Kim J., Park C., Rossi G., Lee S.-M., Gott J. R. I., 2011, Journal of The Korean Astronomical Society, 44, 217

Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2900

King S. F., Merle A., 2012, Journal of Cosmology and Astroparticle Physics, 2012, 016

Knebe A. et al., 2011, Monthly Notices of the Royal Astronomical Society, 415, 2293

Knobel C. et al., 2012, The Astrophysical Journal, 755, 48

Knobel C., Lilly S. J., Woo J., Kovač K., 2015, The Astrophysical Journal, 800, 24

Komatsu E. et al., 2011, The Astrophysical Journal Supplement Series, 192, 18

Kormendy J., Ho L. C., 2013, Annual Review of Astronomy and Astrophysics, 51, 511

Kowal C. T., 1968, The Astronomical Journal, 73, 1021

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlober S., Allgood B., Primack J. R., 2004, The Astrophysical Journal, 609, 35

Lacey C. G. et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3854

Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, Monthly Notices of the Royal Astronomical Society, 251, 128

Laigle C. et al., 2016, The Astrophysical Journal Supplement Series, 224, 24

Landy S. D., Szalay A. S., 1993, The Astrophysical Journal, 412, 64

Larson R., 1974, Monthly Notices of the Royal Astronomical Society, 169, 229

Lawrence A. et al., 2007, Monthly Notices of the Royal Astronomical Society, 379, 1599

Le Borgne J.-F. et al., 2003, Astronomy & Astrophysics, 402, 433

Le Fèvre O. et al., 2013, Astronomy & Astrophysics, 559, A14

Le Fèvre O. et al., 2015, Astronomy & Astrophysics, 576, A79

Lee B. et al., 2013, The Astrophysical Journal, 774, 47

Lesgourgues J., Pastor S., 2014, New Journal of Physics, Volume 16, Issue 6, article id. 065002 (2014)., 16

Lewis I. et al., 2002, Monthly Notices of the Royal Astronomical Society, 334, 673

Li C. et al., 2012, Monthly Notices of the Royal Astronomical Society, 419, 1557

Liddle A., 1999, High Energy Physics and Cosmology, 1998 Summer School, ICTP, Trieste, Italy, 29 June - 17 July 1998, 260

Liddle A. R., Lyth D. H. D. H., 2000, Cosmological inflation and large-scale structure. Cambridge University Press, p. 400

Limber D. N., 1954, The Astrophysical Journal, 119, 655

Lindsay S. N., Jarvis M. J., McAlpine K., 2014, Monthly Notices of the Royal Astronomical Society, 440, 2322

Ling E. N., Barrow J. D., Frenk C. S., 1986, Monthly Notices of the Royal Astronomical Society, 223, 21P

Lu Y., Mo H. J., Weinberg M. D., Katz N., 2011, Monthly Notices of the Royal Astronomical Society, 416, 1949

Madau P., Dickinson M., 2014, Annual Review of Astronomy and Astrophysics, 52, 415

Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, Monthly Notices of the Royal Astronomical Society, 283, 1388

Madau P., Pozzetti L., Dickinson M., 1998, The Astrophysical Journal, 498, 106

Mambrini Y., Toma T., 2015, The European Physical Journal C, 75, 570

Mandelbaum R. et al., 2014, The Astrophysical Journal Supplement Series, 212, 5

Mann R. G., Peacock J. A., Heavens A. F., 1998, Monthly Notices of the Royal Astronomical Society, 293, 209

McAlpine K., Smith D. J. B., Jarvis M. J., Bonfield D. G., Fleuren S., 2012, Monthly Notices of the Royal Astronomical Society, 423, 132

McAlpine S. et al., 2016, Astronomy and Computing, 15, 72

McCracken H. J. et al., 2012, Astronomy & Astrophysics, 544, A156

McCracken H. J. et al., 2007, The Astrophysical Journal Supplement Series, 172, 314

McCracken H. J. et al., 2015, Monthly Notices of the Royal Astronomical Society, 449, 901

McGaugh S. S., 2012, The Astronomical Journal, 143, 40

McGee S. L., Bower R. G., Balogh M. L., 2014a, Monthly Notices of the Royal Astronomical Society: Letters, 442, L105

McGee S. L., Goto R., Balogh M. L., 2014b, Monthly Notices of the Royal Astronomical Society, 438, 3188

McLeod D. J., McLure R. J., Dunlop J. S., 2016, Monthly Notices of the Royal Astronomical Society, 459, 3812

McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, Monthly Notices of the Royal Astronomical Society, 395, 2196

McLure R. J. et al., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2696

McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, Monthly Notices of the Royal Astronomical Society, 381, 75

Meneux B. et al., 2009, Astronomy and Astrophysics, 505, 463

Merritt D., Graham A. W., Moore B., Diemand J., Terzić B., 2006, The Astronomical Journal, 132, 2685

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, The Journal of Chemical Physics, 21, 1087

Milgrom M., 1983, The Astrophysical Journal, 270, 365

Mitsou V. A., 2015, Journal of Physics: Conference Series, 651, 012023

Miyazaki S. et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, Proceedings of the SPIE, McLean I. S., Ramsay S. K., Takami H., eds., p. 84460Z

Mo H. J., Jing Y. P., White S. D. M., 1997, Monthly Notices of the Royal Astronomical Society, 284, 189

Mohr J. J. et al., 2008, in Observatory Operations: Strategies, Processes, and Systems II, Proceedings of the SPIE, Brissenden R. J., Silva D. R., eds., p. 70160L

Moore B., Lake G., Katz N., 1998, The Astrophysical Journal, 495, 139

Moster B. P., Naab T., White S. D. M., 2013, Monthly Notices of the Royal Astronomical Society, 428, 3121

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, The Astrophysical Journal, 710, 903

Murray S. G., Power C., Robotham A. S. G., 2013, Monthly Notices of the Royal Astronomical Society: Letters, 434, L61

Mutch S. J., Croton D. J., Poole G. B., 2013, Monthly Notices of the Royal Astronomical Society, 435, 2445

Nantais J. B. et al., 2017, Monthly Notices of the Royal Astronomical Society: Letters, 465, L104

Natarajan A., Yoshida N., 2014, Progress of Theoretical and Experimental Physics, 6B112

Natarajan P. et al., 2017, Monthly Notices of the Royal Astronomical Society

Navarro J. F., Frenk C. S., White S. D. M., 1996, The Astrophysical Journal, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, The Astrophysical Journal, 490, 493

Neistein E., Weinmann S. M., 2010, Monthly Notices of the Royal Astronomical Society, 405, 2717

Newman J. A., Davis M., 2002, The Astrophysical Journal, 564, 567

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, Monthly Notices of the Royal Astronomical Society, 396, 19

Norberg P. et al., 2002, Monthly Notices of the Royal Astronomical Society, 332, 827

Oemler, Augustus J., 1974, The Astrophysical Journal, 194, 1

Oesch P. A. et al., 2016, The Astrophysical Journal, 819, 129

Oke J. B., Gunn J. E., 1983, The Astrophysical Journal, 266, 713

Oknyanskij V. L., 1999, Odessa Astronomical Publications, 12, 99

Ouchi M. et al., 2009, The Astrophysical Journal, 706, 1136

Ouchi M. et al., 2010, The Astrophysical Journal, 723, 869

Pahwa I., Paranjape A., 2016, eprint arXiv:1612.00464

Paranjape A., Kovač K., Hartley W. G., Pahwa I., 2015, Monthly Notices of the Royal Astronomical Society, 454, 3030

Parry W., 1977, Physics Letters A, 60, 265

Parzen E., 1962, The Annals of Mathematical Statistics, 33, 1065

Patel S. G., Holden B. P., Kelson D. D., Illingworth G. D., Franx M., 2009, The Astrophysical Journal, 705, L67

Peacock J. A., Smith R. E., 2000, Monthly Notices of the Royal Astronomical Society, 318, 1144

Peacock . et al., 2001, Nature, 410, 169

Peebles P. J. E., 1980, The Large Scale Structure of the Universe. Princeton University Press

Peebles P. J. E., 1993, Principles of physical cosmology. Princeton University Press

Peebles P. J. E., Yu J. T., 1970, The Astrophysical Journal, 162, 815

Peirani S. et al., 2016, eprint arXiv:1611.09922

Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192

Peng Y.-j. et al., 2010, The Astrophysical Journal, 721, 193

Pentericci L. et al., 2014, The Astrophysical Journal, 793, 113

Penzias A. A., Wilson R. W., 1965, The Astrophysical Journal, 142, 419

Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, Monthly Notices of the Royal Astronomical Society, 381, 1053

Percival W. J. et al., 2010, Monthly Notices of the Royal Astronomical Society, 401, 2148

Perlmutter S. et al., 1999, The Astrophysical Journal, 517, 565

Prescott M. et al., 2011, Monthly Notices of the Royal Astronomical Society, 417, 1374

Press W. H., Schechter P., 1974, The Astrophysical Journal, 187, 425

Primack J. R., 2001, eprint arXiv:astro-ph/0112336

Rawlings S., Jarvis M. J., 2004, Monthly Notices of the Royal Astronomical Society, 355, L9

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, The Astrophysical Journal, 771, 30

Rees M. J., Ostriker J. P., 1977, Monthly Notices of the Royal Astronomical Society, 179, 541

Rieke G. H., Lebofsky M. J., Thompson R. I., Low F. J., Tokunaga A. T., 1980, The Astrophysical Journal, 238, 24

Riess A. G. et al., 1998, The Astronomical Journal, 116, 1009

Risaliti G., Elvis M., 2004, in Supermassive Black Holes in the Distant Universe, Kluwer Academic Publishers, pp. 187–224

Roberts M. S., Haynes M. P., 1994, Annual Review of Astronomy and Astrophysics, 32, 115

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, The Astrophysical Journal, 802, L19

Robertson H. P., 1935, The Astrophysical Journal, 82, 284

Roche N., Eales S. A., 1999, Monthly Notices of the Royal Astronomical Society, 307, 703

Rodighiero G. et al., 2010, Astronomy and Astrophysics, 518, L25

Rodriguez F., Merchán M., Sgró M. A., 2015, Astronomy & Astrophysics, 580, A86

Rosenblatt M., 1956, The Annals of Mathematical Statistics, 27, 832

Röttgering H. et al., 2011, Journal of Astrophysics and Astronomy, 32, 557

Rubin V. C., Thonnard N., Ford, W. K. J., 1980, The Astrophysical Journal, 238, 471

Ryle M., Clarke R. W., 1961, Monthly Notices of the Royal Astronomical Society, 122, 349

Sachs R. K., Wolfe A. M., 1967, The Astrophysical Journal, 147, 73

Saghiha H., Simon P., Schneider P., Hilbert S., 2017, Astronomy & Astrophysics

Saintonge A., Masters K. L., Marinoni C., Spekkens K., Giovanelli R., Haynes M. P., 2008, Astronomy and Astrophysics, 478, 57

Santos M. G., Cooray A., Knox L., 2005, The Astrophysical Journal, 625, 575

Sarazin C. L., 1986, Reviews of Modern Physics, 58, 1

Sawala T. et al., 2015, Monthly Notices of the Royal Astronomical Society, 448, 2941

Schawinski K. et al., 2014, Monthly Notices of the Royal Astronomical Society, 440, 889

Schaye J. et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521

Schechter P., 1976, The Astrophysical Journal, 203, 297

Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, The Astrophysical Journal, 546, 20

Scott D., Smoot G., 2006, eprint arXiv:astro-ph/0601307

Scoville N. et al., 2007, The Astrophysical Journal Supplement Series, 172, 38

Sefusatti E., Crocce M., Pueblas S., Scoccimarro R., 2006, Physical Review D, 74, 023522

Seljak U., 2000, Monthly Notices of the Royal Astronomical Society, 318, 203

Seljak U., 2009, Physical Review Letters, 102, 021302

Shapley A. E., 2011, Annual Review of Astronomy and Astrophysics, 49, 525

Sharma M., Theuns T., Frenk C., Bower R., Crain R., Schaller M., Schaye J., 2016, Monthly Notices of the Royal Astronomical Society: Letters, 458, L94

Sherwin B. D. et al., 2012, Physical Review D, 86, 083006

Sheth R. K., Mo H. J., Tormen G., 2001, Monthly Notices of the Royal Astronomical Society, 323, 1

Sheth R. K., Tormen G., 1999, Monthly Notices of the Royal Astronomical Society, 308, 119

Silk J., 2010, Proceedings of the International Astronomical Union, 6, 273

Silk J., 2013, The Astrophysical Journal, 772, 112

Silk J., Mamon G. A., 2012, Research in Astronomy and Astrophysics, 12, 917

Silk J., Norman C., 2009, The Astrophysical Journal, 700, 262

Simon P., Hetterscheidt M., Wolf C., Meisenheimer K., Hildebrandt H., Schneider P., Schirmer M., Erben T., 2009, Monthly Notices of the Royal Astronomical Society, 398, 807

Skibba R. A. et al., 2009, Monthly Notices of the Royal Astronomical Society, 399, 966

Smethurst R. J. et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 435

Smith D. J. B. et al., 2016, SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics

Smoot G. F., 1999, in Conference on 3K cosmology, ASCE, pp. 1–10

Sofue Y., Rubin V., 2001, Annual Review of Astronomy and Astrophysics, 39, 137

Sohn J., Geller M. J., Zahid H. J., Fabricant D. G., Diaferio A., Rines K. J., 2016, eprint arXiv:1612.06428

Somerville R. S., Davé R., 2015, Annual Review of Astronomy and Astrophysics, 53, 51

Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, The Astrophysical Journal, 600, L171

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

Stark D. P., 2016, Annual Review of Astronomy and Astrophysics, 54, 761

Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, The Astrophysical Journal, Volume 697, Issue 2, pp. 1493-1511 (2009)., 697, 1493

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, The Astrophysical Journal, 462, L17

Steidel C. C., Hamilton D., 1992, The Astronomical Journal, 104, 941

Steinhardt C. L., Capak P., Masters D., Speagle J. S., 2016, The Astrophysical Journal, 824, 21

Steinhauser D., Schindler S., Springel V., 2016, Astronomy & Astrophysics, 591, A51

Strateva I. et al., 2001, The Astronomical Journal, 122, 1861

Stringer M. J., Benson A. J., Bundy K., Ellis R. S., Quetin E. L., 2009, Monthly Notices of the Royal Astronomical Society, 393, 1127

SubbaRao M. U., Aragón-Calvo M. A., Chen H. W., Quashnock J. M., Szalay A. S., York D. G., 2008, New Journal of Physics, 10, 125015

Sunyaev R. A., Zeldovich Y. B., 1970, Astrophysics and Space Science, 7, 3

Suzuki N. et al., 2012, The Astrophysical Journal, 746, 85

Szapudi I., Szalay A. S., 1998, The Astrophysical Journal, 494, L41

Tammann G. A., 1979, Scientific research with the Space Telescope (IAU Colloq. No. 54), 2111, 263

Taylor M. B., 2005, Astronomical Data Analysis Software and Systems XIV ASP Conference Series, 347

Tegmark M. et al., 2004, The Astrophysical Journal, 606, 702

Teyssier R., 2002, Astronomy and Astrophysics, 385, 337

Thornton R. J. et al., 2016, The Astrophysical Journal Supplement Series, 227, 21

Tiley A. L. et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 103

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, The Astrophysical Journal, 688, 709

Tinker J. L., Hahn C., Mao Y.-Y., Wetzel A. R., Conroy C., 2017, eprint arXiv:1702.01121

Tinker J. L., Leauthaud A., Bundy K., George M. R., Behroozi P., Massey R., Rhodes J., Wechsler R. H., 2013, The Astrophysical Journal, 778, 93

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

Tisserand P. et al., 2007, Astronomy and Astrophysics, 469, 387

Tomczak A. R. et al., 2014, The Astrophysical Journal, 783, 85

Tonnesen S., Cen R., 2014, The Astrophysical Journal, 788, 133

Trenti M., Stiavelli M., 2008, The Astrophysical Journal, 676, 767

Tully R. B., Fisher J. R., 1977, Astronomy and Astrophysics, 54, 661

Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, Astronomy and Astrophysics, 506, 647

Vale A., Ostriker J. P., 2006, Monthly Notices of the Royal Astronomical Society, 371, 1173

van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, Monthly Notices of the Royal Astronomical Society, 430, 725

Vernstrom T. et al., 2014, Monthly Notices of the Royal Astronomical Society, 440, 2791

Viola M. et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 3529

Vogelsberger M. et al., 2014, Nature, 509, 177

Wang L. et al., 2016, eprint arXiv:1612.04540

Wang L., Li C., Kauffmann G., De Lucia G., 2006, Monthly Notices of the Royal Astronomical Society, 371, 537

Wang Q. D., 2011, Proceedings of the International Astronomical Union, 7, 183

Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, The Astrophysical Journal, 646, 881

Watson D., Denney K. D., Vestergaard M., Davis T. M., 2011, The Astrophysical Journal, 740, L49

Weinberg D. H., Dave R., Katz N., Hernquist L., 2004, The Astrophysical Journal, 601, 1

Weinberg N. N., Kamionkowski M., 2003, Monthly Notices of the Royal Astronomical Society, 341, 251

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, Monthly Notices of the Royal Astronomical Society, 366, 2

Welikala N. et al., 2015, Monthly Notices of the Royal Astronomical Society, 455, 1629

White S. D. M., Rees M. J., 1978, Monthly Notices of the Royal Astronomical Society, 183, 341

White S. V., Jarvis M. J., Haussler B., Maddox N., 2015, Monthly Notices of the Royal Astronomical Society, 448, 2665

Willett K. W. et al., 2013, Monthly Notices of the Royal Astronomical Society, 435, 2835

Williams W. L. et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 2385

Wilman R. J. et al., 2008, Monthly Notices of the Royal Astronomical Society, 388, 1335

Wojtak R., Hansen S. H., Hjorth J., 2011, Nature, 477, 567

Woo J. et al., 2013, Monthly Notices of the Royal Astronomical Society, 428, 3306

Wu X.-P., Chiueh T., Fang L.-Z., Xue Y.-J., 1998, Monthly Notices of the Royal Astronomical Society, 301, 861

Yang X., Mo H. J., Bosch F. C. v. d., 2003, Monthly Notices of the Royal Astronomical Society, 339, 1057

Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, Monthly Notices of the Royal Astronomical Society, 358, 217

Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, The Astrophysical Journal, 671, 153

Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, The Astrophysical Journal, 752, 41

Yoshikawa K., Taruya A., Jing Y. P., Suto Y., 2001, The Astrophysical Journal, 558, 520

Zaroubi S., 2013, in The First Galaxies, Astrophysics and Space Science Library, Vol. 396, Springer, pp. 45–101

Zehavi I. et al., 2005, The Astrophysical Journal, 621, 22

Zehavi I. et al., 2011, The Astrophysical Journal, 736, 59

Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005, The Astrophysical Journal, 624, 505

Zheng Z. et al., 2005, The Astrophysical Journal, 633, 791

Zinn P.-C., Middelberg E., Norris R. P., Dettmar R.-J., 2013, The Astrophysical Journal, 774, 66

Zu Y., Mandelbaum R., 2015, Monthly Notices of the Royal Astronomical Society, 454, 1161

Zu Y., Mandelbaum R., 2016, Monthly Notices of the Royal Astronomical Society, 457, 4360

Zwart J. T. L., Jarvis M. J., Deane R. P., Bonfield D. G., Knowles K., Madhanpall N., Rahmani H., Smith D. J. B., 2014, Monthly Notices of the Royal Astronomical Society, 439, 1459

Zwart J. T. L., Santos M., Jarvis M. J., 2015, Monthly Notices of the Royal Astronomical Society, 453, 1740

Zwicky F., 1937, The Astrophysical Journal, 86, 217