Galaxies and their Host Dark Matter Structures

by

ChangHoon Hahn

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Physics

New York University

May 2017

Professor Michael R. Blanton

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Dedication

To my best friend and partner, Lidia

Acknowledgements

First and foremost, I would like to thank my advisor Michael R. Blanton. He very patiently guided me when I started graduate school and later gave me the freedom to explore on my own while always being available for guidance and wisdom when I needed it. I am extremely thankful for his amazing patience, encouragement, and thoughtful mentoring.

I am deeply grateful to Roman Scoccimarro for his patient mentoring, his deep scientific intuitions, and his insightful yet pedagogical explanations whenever I stepped into his office. I am also very thankful to Jeremy L. Tinker for his warm mentoring, constant encouragement and support, and always having his door open for scientific discussions. I would also like to thank David Hogg for his guidance, support, and unending energetic enthusiasm for science (even fiber collisions).

This dissertation and my graduate school career would not have been possible without the help of Dan Foreman-Mackey, Boris Leistedt, Maryam Modjaz, John Moustakas, Khee-Gan Lee, Fransisco Prada, Sergio Rodríguez-Torres, Shenglong Wang, Andrew Wetzel, Renbin Yan, Guangtun Zhu, and many others.

I would like to thank my classmates at NYU: Ruben Gepner, Mark Hannel, Oliver Janssen, Doycho Karagyozov, Alex Malz, Nityasri Mandyam, Christine Middleton, Geoffrey Ryan, Cato Sanford, Rob Speare, Mohammadjavad Vakili, Kilian Walsh, and many others. In addition to the intellectually stimulating environment and craic they provided, their friendships made my graduate school experience truly enjoyable. Outside of the department, I would also like to thank Adam, Eugene, Greg, Jasmine, Shyr, and Vinny for all the good times in NYC.

Finally, I want to thank Lidia Cano, to whom this thesis is dedicated, for the unending love and support she has given me over the years. She has taught me so much about life. Without her by my side, I would not be where I am today.

Abstract

Through their connection with dark matter structures, galaxies act as tracers of the underlying matter distribution in the Universe. Their observed spatial distribution allows us to precisely measure large scale structure and effectively test cosmological models that explain the content, geometry, and history of the Universe. Current observations from galaxy surveys such as the Baryon Oscillation Spectroscopic Survey have already probed vast cosmic volumes with millions of galaxies and ushered in an era of precision cosmology. The next surveys will probe volumes over an order of magnitude larger. With this unprecedented statistical power, the bottleneck of scientific discovery is in the methodology.

In this dissertation, I address major methodological challenges in constraining cosmology with the large-scale spatial distribution of galaxies. I develop a robust framework for treating systematic effects, which significantly bias galaxy clustering measurements. I apply new innovative approaches to probabilistic parameter inference that challenge and test incorrect assumptions of the standard approach. Furthermore, I use precise predictions of structure formation from cosmology and observations of galaxies during the last eight billion years to develop detailed models of how galaxies are impacted by their host dark matter structures. These models provide key insight into the galaxy-halo connection, which bridges the gap between cosmology theory and observations. They also answer crucial questions of how galaxies form and evolve. The developments in this dissertation will help unlock the full potential of future observations and allow us to precisely test cosmological models, General Relativity and modified gravity scenarios, and even particle physics theory beyond the Standard Model.

Contents

Dedication . . . iii

Acknowledgements . . . iv

Abstract . . . v

List of Figures . . . ix

List of Tables . . . xxx

List of Appendices . . . xxxi

Introduction **1**

0.1 Large Scale Structure in ΛCDM . . . 2

0.2 Redshift-Space Distortions . . . 8

0.3 Weighting Neutrinos with Galaxies . . . 11

0.4 Analyzing Galaxy Clustering . . . 13

1 The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles **22**

1.1 Chapter Abstract . . . 22

1.2 Introduction . . . 23

1.3 Fiber-collided Mock catalogs . . . 27

1.4 Fiber Collision Methods . . . 35

1.5 Summary and Conclusions . . . 67

2 Approximate Bayesian Computation in Large Scale Structure: constraining the galaxy-halo connection 72

2.1 Chapter Abstract 72

2.2 Introduction 73

2.3 Methods 77

2.4 ABC at work 87

2.5 Summary and Conclusion 102

3 PRIMUS: Effects of Galaxy Environment on the Quiescent Fraction Evolution at $z < 0.8$ 112

3.1 Chapter Abstract 112

3.2 Introduction 113

3.3 Sample Selection 116

3.4 Results: Stellar Mass Function 130

3.5 Results: Quiescent Fraction 135

3.6 Summary and Discussion 149

4 Star Formation Quenching Timescale of Central Galaxies in a Hierarchical Universe 153

4.1 Chapter Abstract 153

4.2 Introduction 154

4.3 Central Galaxies of SDSS DR7 157

4.4 Simulated Central Galaxy Catalog 158

4.5 Star Formation in Central Galaxies 163

4.6 Results 182

4.7 Discussion 188

4.8 Summary 196

Conclusion **199**

Appendices **201**

Bibliography **209**

List of Figures

1 Schematic diagram that illustrates the evolution of the density fluctuations in the early Universe through inflation, radiation-dominated epoch, matter-radiation equality, and matter-dominated epoch. The evolution of the Hubble radius (solid line) remains flat during inflation (flat), scales by $\propto a^2$ during radiation domination, and $\propto a^{3/2}$ during matter domination. a_{eq} marks matter radiation equality. The physical lengths of three constant comoving scales are marked by dashed, dotted, and dot-dashed lines. The dashed line represents physical lengths of perturbations that enter the horizon during matter-radiation equality $\lambda_{phys} = \lambda_{eq}$. The dot-dashed line mark perturbations that enter the horizon during radiation domination with $\lambda_{phys} < \lambda_{eq}$. The dotted line mark perturbations that enter the horizon during matter domination with $\lambda_{phys} > \lambda_{eq}$. As described in the text, the growth of perturbations with $\lambda_{phys} < \lambda_{eq}$ are suppressed because they enter the horizon during the radiation dominated era. The evolution of the density perturbation through these epochs are quantified through the transfer function $T(k)$. 6

2 Amplitude of the reduced galaxy bispectrum $Q(k_1, k_2, k_3)$ plotted as a function of ratios k_2/k_1 and k_3/k_1, which describe the triangle configurations. The $Q(k_1, k_2, k_3)$ in the left panel is calculated from a perturbation theory model while the right panel presents $Q(k_1, k_2, k_3)$ of BOSS Data Release 12 CMASS galaxy sample using the Scoccimarro (2015) estimator. 17

1.1 Normalized galaxy redshift distribution of the Nseries (orange), QPM (blue), and BigMultiDark (red) mock catalogs. The normalized redshift distribution of BOSS DR12 CMASS sample galaxies is also plotted (black). Each of the distributions were computed with a bin size of $\Delta z = 0.025$. All of the mock catalogs used in this work closely trace the BOSS CMASS redshift distribution. 29

1.2 Power spectrum monopole $P_0(k)$ and quadrupole $|P_2(k)|$ measurements for the Nseries (orange), QPM (blue), and BigMultiDark (red) mock catalogs (Section 1.3). The $P_l(k)$ measurements for the Nseries and QPM mock catalogs are averaged over the multiple mock realizations and the width of the power spectra represents the sample variance ($\sigma_l(k)$; Eq. 1.10) of the realizations. For the quadrupole, we plot the $|P_2(k)|$ instead of $P_2(k)$ because the measurement becomes negative for $k \gtrsim 0.35$ h/Mpc. For comparison, we also include the monopole and quadrupole power spectra of the BOSS DR12 CMASS sample, which are calculated using the same estimator but with statistical weights described in Eq. (1.9). While fiber collisions are inevitably included in the BOSS CMASS power spectra, they are *not* yet applied to the mock catalogs power spectra measurements above. 32

1.3 The fiber collision power spectrum residual, $(P_l^{\mathrm{NN}} - P_l^{\mathrm{true}})$ (Section 1.4.1), for the monopole (top), quadrupole (middle), and hexadecapole (bottom) of the Nseries (left), QPM (middle), and BigMultiDark (right) mock catalogs. For the Nseries and QPM mocks, we plot the sample variances $\sigma_l(k)$ (grey shaded region) of $P_l^{\mathrm{true}}(k)$ for comparison. The power spectrum residual for the NN method is an improvement over the residual with no correction $(\Delta P_l^{\mathrm{NoW}}(k)$; x) at most scales probed. However, we highlight that at $k > 0.1$ h/Mpc and $k > 0.2$ h/Mpc, for the monopole and quadrupole respectively, the residuals from fiber collision surpass the sample variance. At smaller scales, NN method does not sufficiently account for the effects of fiber collisions in $P_l(k)$ measurements. 36

1.4 *Top Panel*: The normalized residuals, $1-\overline{P_0^{\mathrm{NN}}}/\overline{P_0^{\mathrm{true}}}(k)$, of the NN method for the Nseries (orange), QPM (blue), and BigMultiDark (red) power spectrum monopole. We also plot the normalized sample variance $\sigma_0(k)/P_0(k)$ (gray shaded region) of the Nseries mocks for comparison. The QPM $\sigma_0(k)/P_0(k)$ is effectively the same as the Nseries $\sigma_0(k)/P_0(k)$, so we do not included in the figure. The comparison reveals that the effect of fiber collisions not only biases the power spectrum beyond sample variance at $k \gtrsim 0.1\ h/\mathrm{Mpc}$, but that the effect increases relative to sample variance at smaller scales. At $k = 0.2\ h/\mathrm{Mpc}$, the normalized residual is greater than 4 times the normalized sample variance. *Bottom Panel*: We mark k_{χ^2} where $\Delta\chi^2(k_{\chi^2}) = 1$ (Eq. 1.11) for the NN method. $k_{\chi^2}^{\mathrm{NN}}$ is a conservative scale limit of the NN method. Arrows above the dashed line mark k_{χ^2} for the monopole while the arrows below the dashed line mark k_{χ^2} for the quadrupole. The color of the arrows indicate the mock catalog: Nseries (orange), QPM (blue), and BigMultiDark (red). Averaged over the three mock catalogs, we get $k_{\chi^2}^{\mathrm{NN}} = 0.068$ and $0.17\ h/\mathrm{Mpc}$. for the monopole and quadrupole respectively. 37

1.5 Normalized distribution of d_{LOS} for Nseries (orange), QPM (blue), and BigMultiDark (green) mock catalogs. The normalized d_{LOS} distribution of BOSS DR12 is also plotted (black). The mock catalog distributions have bin sizes of $\Delta d = 0.2\,\mathrm{Mpc}$, while the CMASS distribution has a bin size of $\Delta d = 0.5\,\mathrm{Mpc}$. The distribution extends beyond the range of the above plot to $\sim \pm 500$ Mpc. In the discussion of Section 1.4.2, we focus mainly on the peak of the distribution at roughly $-20\ \mathrm{Mpc} < d_{\mathrm{LOS}} < 20\ \mathrm{Mpc}$. 42

1.6 The power spectrum residual of the line-of-sight reconstruction (LRec) method (Section 1.4.2), $\Delta P_\ell \equiv P_l^{\mathrm{LRec}} - P_l^{\mathrm{true}}$, for the monopole (top) and quadrupole (bottom) power spectra of the Nseries (left), QPM (middle), and BigMultiDark (right) mock catalogs. We again plot the Nseries and QPM sample variances, $\sigma_l(k)$. The residuals for the monopole show good agreement between P_0^{LRec} and P_0^{true} for the entire k range. For the quadrupole, while the LOS Reconstruction method improves the residuals compared to the NN method at small scales ($k > 0.2\ h/\mathrm{Mpc}$), the residuals remain comparable to sample variance at $k = 0.2\ h/\mathrm{Mpc}$. In the top panels, we include the residuals from the fiber collision correction method of Gil-Marín et al. (2014) (dashed). As the Gil-Marín et al. (2014) method supplements the NN method with adjustments to the constant shot noise term of the estimator, it fails to correct for the k dependence of the effect and is insufficient in accounting for fiber collisions at small scales. We do not include the correction method of Beutler et al. (2014b) because they marginalize over a constant stochasticity term in their analysis so the effect of their correction on $P(k)$ is not straightforward. 49

1.7 *Top Panel*: The normalized residual, $1-P_l^{\mathrm{LRec}}/P_l^{\mathrm{true}}$, for the Nseries (orange), QPM (blue), and BigMultiDark (red) monopole power spectra. The normalized sample variance $\sigma_l/P_l(k)$ (gray shaded region) of the Nseries mocks is plotted for comparison. At $k = 0.1\ h/\mathrm{Mpc}$, where the NN method residuals exceeds sample variance, the average normalized residual for the LRec method is 0.25% compared to 1.5% normalized sample variance. In fact, the average residual stays below the sample variance until $k = 0.53\ h/\mathrm{Mpc}$. *Bottom Panel*: We mark $k_{\chi^2}^{\mathrm{LRec}}$ for the monopole (arrows above the dashed line) and quadrupole (arrows below the dashed line). The average $k_{\chi^2}^{\mathrm{LRec}}$ for the mock catalogs are 0.29 and 0.14 h/Mpc for the monopole and quadrupole respectively. For comparison, we mark $k_{\chi^2}^{\mathrm{NN}}$ (black) from Section 1.4.1. We also include k_{χ^2} of the Gil-Marín et al. (2014) correction method (gray) for the monopole. The LOS reconstruction method significantly extends k_{χ^2} beyond that of the NN method and Gil-Marín et al. (2014) for $l = 0$. However, it does not improve k_{χ^2} for the quadrupole. 50

1.8 $1-(1+\xi^{\mathrm{NN}})/(1+\xi^{\mathrm{true}})$ as a function of transverse displacement, r_p, and line-of-sight displacement π (left). The color bar represents the value of this quantity. Note there is no detectable dependence on π. The dashed vertical line (black) represents the constant $r_p = D_{\mathrm{fc}}(z = 0.55)$ (Section 1.4.3). We also plot $1-(1+\xi^{\mathrm{NN}})/(1+\xi^{\mathrm{true}})$ projected along π (right). In the left panel, the $r_p = D_{\mathrm{fc}}(z = 0.55)$ vertical line and the sharp cut-off of the contour show good agreement with the expected characteristic scale. In the right panel, the projected $1-(1+\xi^{\mathrm{NN}})/(1+\xi^{\mathrm{true}})$ is in good agreement with $f_s W_{\mathrm{fc}}(r_p)$. The agreement in both panels justify the characterization of the effect of fiber collisions on the 2PCF in Eq. (1.21). 55

1.9 Comparison of the power spectrum residuals from NN-corrected fiber collisions $\Delta P_l = P_l^{\mathrm{NN}} - P_l^{\mathrm{true}}$ (dashed black) with the ΔP_l from the effective window method obtained by adding Eqs. (1.26) and (1.34) (orange) for the monopole (left) and quadrupole (right). The standard deviation of the power spectrum residual, $\sigma_{\Delta P_l}$, for the Nseries mock catalogs is shaded in gray. 59

1.10 Comparison of the correlated power spectrum residuals from unreliable modes obtained from mocks (dashed), Eq. (1.38), to the polynomial approximation of Eq. (1.37) for $l' \leq 18$ (orange). The left and right panels correspond to $l = 0$ and 2 respectively. The gray shaded region is the standard deviation for the Nseries ($P_l^{\mathrm{NN}} - P_l^{\mathrm{true}}$). We also include Eq. (1.37) evaluated only for $l' \leq 2$ (blue). The agreement between Eq. (1.37) for $l' \leq 2$ and Eq. (1.38) demonstrate that while higher orders of l' are necessary to properly model $\Delta P_l^(k)$ at higher k values, for $k < k_{\mathrm{trust}}$ (0.3 h/Mpc above) $l' \leq 2$ are sufficient. 63

2.1 The two-point correlation function $\xi_{\mathrm{gg}}(r)$ (left) and group multiplicity function $\zeta_{\mathrm{g}}(N)$ (right) summary statistics of the mock observations generated from the "true" HOD parameters described in Section 2.4.1. The width of the shaded region corresponds to the square root of the covariance matrix diagonal elements (Eq. 2.14). In our ABC analysis, we treat the $\xi_{\mathrm{gg}}(r)$ and $\zeta_{\mathrm{g}}(N)$ above as the summary statistics of the observation. 88

2.2 We demonstrate the evolution of the ABC particles, $\vec{\theta}_t$, over iterations $t = 1$ to 9 in the $\log \mathcal{M}_{min}$ and $\log \mathcal{M}_1$ parameter space. $\bar{n}$ and $\zeta_{\rm g}(N)$ are used as observables for the above results. For reference, in each panel, we include the "true" HOD parameters (black star) listed in Section 2.4.1. The initial distance threshold, $\vec{\epsilon}_1 = [\infty, \infty]$ at $t = 1$ (top left) so the $\vec{\theta}_1$ spans the entire range of the prior distribution, which is also the range of the panels. We see for $t < 5$, the parameter space occupied by the ABC $\vec{\theta}_t$ shrinks dramatically. Eventually when the algorithm converges, $t > 7$, the parameter space occupied by $\vec{\theta}_t$ no longer shrinks and their distributions represent the posterior distribution of the parameters. At $t = 9$, the final iteration, the ABC algorithm has converged and we find that $\vec{\theta}_{\rm true}$ lies safely within the 68% confidence region. 89

2.3 We illustrate the convergence of the ABC algorithm through the evolution of the ABC particle distribution as a function of iteration for parameters $\log \mathcal{M}_{\rm min}$ (left), α (center), and $\log \mathcal{M}_1$ (right). The top panel corresponds our ABC results using the observables $(\bar{n}, \zeta_{\rm g}(N))$, while the lower panel plots corresponds to the ABC results using $(\bar{n}, \xi_{\rm gg}(r))$. The distributions of parameters show no significant change after $t > 7$, which suggests that the ABC algorithm has converged. 92

2.4 We present the constraints on the Zheng et al. (2007a) HOD model parameters obtained from our ABC-PMC analysis using $\bar{n}$ and $\xi_{\mathrm{gg}}(r)$ as observables. The diagonal panels plot the posterior distribution of each HOD parameter with vertical dashed lines marking the 50% quantile and 68% confidence intervals of the distribution. The off-diagonal panels plot the degeneracies between parameter pairs. The range of each panel corresponds to the range of our prior choice. The "true" HOD parameters, listed in Section 2.4.1, are also plotted in each of the panels (black). For $\log \mathcal{M}_0$, α, and $\sigma_{\log M}$, the "true" parameter values lie near the center of the 68% confidence interval of the posterior distribution. For $\log \mathcal{M}_1$ and $\log \mathcal{M}_{\mathrm{min}}$, which have tight constraints, the "true" values lie within the 68% confidence interval. Ultimately, the ABC parameter constraints we obtain in our analysis are consistent with the "true" HOD parameters. 106

2.5 Same as Figure 2.4 but for our ABC analysis using $\bar{n}$ and $\zeta_{\mathrm{g}}(N)$ as observables. The ABC parameter constraints we obtain are consistent with the "true" HOD parameters. 107

2.6 We compare the ABC-PMC posterior prediction for the observables $\xi_{\rm gg}(r)$ (left) and $\zeta_{\rm g}(N)$ (right) (orange; Section 2.4.3) to $\xi_{\rm gg}(r)$ and $\zeta_{\rm g}(N)$ of the mock observation (black) in the top panels. In the lower panels, we plot the ratio between the ABC-PMC posterior predictions for $\xi_{\rm gg}$ and $\zeta_{\rm g}$ to the mock observation $\xi_{\rm gg}^{\rm obvs}$ and $\zeta_{\rm g}^{\rm obvs}$. The darker and lighter shaded regions represent the 68% and 95% confidence regions of the posterior predictions, respectively. The error-bars represent the square root of the diagonal elements of the error covariance matrix (equation 2.14) of the mock observations. Overall, the observables drawn from the ABC-PMC posteriors are in good agreement with $\xi_{\rm gg}$ and $\zeta_{\rm g}$ of the mock observations. The lower panels demonstrate that for both observables, the error-bars of the mock observations lie within the 68% confidence interval of the ABC-PMC posterior predictions. 108

2.7 We compare the $\log \mathcal{M}_{\rm min}$, α, and $\log \mathcal{M}_1$ parameter constraints from ABC-PMC (orange) to constraints from the Gaussian pseudo-ikelihood MCMC (blue) using $\bar{n}_{\rm g}$ and $\xi_{\rm gg}(r)$ as observables. The *top* panels compares the two methods' marginalized posterior PDFs over the parameters. In the *bottom* panels, we include box plots marking the confidence intervals of the posterior distributions. The boxes represent the 68% confidence interval while the "whiskers" represent the 95% confidence interval. We mark the "true" HOD parameters with vertical black dashed line. The marginalized posterior PDFs obtained from the two methods are consistent with each other. The ABC-PMC and Gaussian pseudo-likelihood constraints are generally consistent for $\log \mathcal{M}_{\rm min}$ and $\log \mathcal{M}_1$. The ABC-PMC constraint for α is slightly less biased and has slightly larger uncertainty then the constraint from Gaussian pseudo-likelihood analysis. 109

2.8 Same as Figure 2.7, but both the ABC-PMC analysis and the Gaussian pseudo-likelihood MCMC analysis use $\bar{n}_{\mathrm{g}}$ and $\zeta_{\mathrm{g}}(N)$ as observables. Both methods derive constraints consistent with the "true" HOD parameters and infer the region of allowed values to similar precision. We note that the MCMC constraint on α is slightly more biased compared to ABC-PMC estimate. This discrepancy may stem from the fact that the use of Gaussian pseudo-likelihood and its associated assumptions is more spurious when modeling the group multiplicity function. 110

2.9 We compare the ABC-PMC (orange) and the Gaussian pseudo-likelihood MCMC (blue) predictions of the 68% and 95% posterior confidence regions over the HOD parameters ($\log \mathcal{M}_{\mathrm{min}}$, α, and $\log \mathcal{M}_1$) using $\bar{n}_{\mathrm{g}}$ and $\xi_{\mathrm{gg}}(r)$ as observables. In each panel, the black star represents the "true" HOD parameters used to generate the mock observations. Both inference methods derive confidence regions consistent with the "true" HOD parameters. 111

2.10 Same as Figure 2.9, but using $\bar{n}_{\mathrm{g}}$ and $\zeta_{\mathrm{g}}(N)$ as observables. Again, the confidence regions derived from both methods are consistent with the "true" HOD parameters used to generate the mock observations. The confidence region of α from the Gaussian pseudo-likelood method is biased compared to the ABC-PMC contours. This may be due to the fact that the true likelihood function that describes $\zeta_{\mathrm{g}}(N)$ deviates significantly from the assumed Gaussian functional form. 111

3.1 Absolute magnitude M_r versus redshift for our mass complete galaxy sample (black squares) with the Environment Defining Population (red circles) plotted on top. Both samples are divided into redshift bins: $z = 0.05 - 0.12$, $0.2 - 0.4$, $0.4 - 0.6$, and $0.6 - 0.8$ (panels left to right). The lowest redshift bin ($z \approx 0.05 - 0.12$; leftmost panel) contain our galaxy sample and EDP selected from SDSS. The rest contain galaxies and EDP selected from PRIMUS. The redshift limits for the lowest redshift bin are empirically selected based on the bright and faint limits of SDSS galaxies. Stellar mass completeness limits, described in Section 3.3.3, are imposed on the galaxy population. Meanwhile, M_r limits are applied to the EDP such that the number density in each panel are equivalent (Section 3.3.5). 117

3.2 Normalized distribution of environment measurements ($n_{\rm env}$) for our mass complete galaxy sample within the survey edges. A fixed cylindrical aperture of $R_{\rm ap} = 2.5$ Mpc and $H_{\rm ap} = 35$ Mpc is used to measure environment. The star-forming galaxies contribution to the distribution is colored in blue and diagonally patterned. The contribution from quiescent galaxies is colored in red. Galaxies with $n_{\rm env} = 0.0$ are in low density environments and galaxies with $n_{\rm env} > 3.0$ are in high density environment. We note that the significant difference among the SDSS distribution and the PRIMUS distributions above is due to the different stellar mass completeness limits imposed on each redshift bin of our galaxy sample. 125

3.3 Evolution of stellar mass functions of star-forming (top) and quiescent (bottom) galaxies in low (left) and high (right) density environments throughout the redshift range $z = 0$–0.8. The environment of each galaxy was calculated using a cylindrical aperture size of $R = 2.5$ Mpc and $H = 35$ Mpc and classified as low environment when $n_{\rm env} = 0.0$ and as high environment when $n_{\rm env} > 3.0$. The SMFs use mass bins of width $\Delta\log(\mathcal{M}/\mathcal{M}_\odot) = 0.2$. In each panel we use shades of blue (star-forming) and orange (quiescent) to represent the SMF at different redshift, higher redshifts being progressively lighter. . . 131

3.4 Evolution of the quiescent fraction $f_{\rm Q}$ for galaxies in low (left) and high (right) density environments for $z < 0.8$. $f_{\rm Q}$s were calculated using the SMFs in Figure 3.3, as described in Section 3.5.1. Darker shading indicates lower redshift and the width represents the standard jackknife uncertainty. 134

3.5 The evolution of the quiescent fraction at fiducial mass, $f_Q(\mathcal{M}_{\rm fid})$, for low (blue) and high (red) density environments within the redshift range $z = 0.0 - 0.8$. We present the $f_Q(\mathcal{M}_{\rm fid})$ evolution for $\mathcal{M}_{\rm fid} = 10^{10.5}\mathcal{M}_\odot$ (solid fill) and $10^{11}\mathcal{M}_\odot$ (patterned fill) with the uncertainty of the best-fit parameter b in Equation 3.7 represented by the width of the line. While the high density $f_{\rm Q}(\mathcal{M}_{\rm fid})$ is greater than low density environment $f_{\rm Q}(\mathcal{M}_{\rm fid})$ over the entire redshift range of our sample, there is a significant increase in $f_{\rm Q}(\mathcal{M}_{\rm fid})$ over cosmic time for both environments. For the environment cut-offs ($n_{\rm env} = 0.0$ for low and $n_{\rm env} > 3.0$ for high), there is no significant difference in the slope of the evolution between the environments. 139

3.6 $f_{\rm Q}(\mathcal{M}_{\rm fid} = 10^{10.5}\mathcal{M}_\odot)$ evolution compared to $f_{\rm red}(\mathcal{M}_* \sim 10^{10.5}\ \mathcal{M}_\odot)$ in the literature: Baldry et al. (2006) (diamond) and Geha et al. (2012a) (cross) from SDSS (panel a), Iovino et al. (2010) (empty square) and Kovač et al. (2014) (triangle) from zCOSMOS (panel b), and Peng et al. (2010) from both SDSS and zCOSMOS (panel c). The $f_{\rm red}$ values from Iovino et al. (2010), Kovač et al. (2014), Baldry et al. (2006), and Peng et al. (2010) are calculated from the best-fit parameterizations presented in the respective works. High density environment is represented in red and low density environment is represented in blue. The $f_{\rm Q}$ value from Geha et al. (2012a) is the $f_{\rm Q}$ value at $\mathcal{M} = 10^{10.55}\mathcal{M}_\odot$. Uncertainties in the Iovino et al. (2010) best-fit $f_{\rm red}$ is omitted due to insufficient information on the cross correlation terms of the fit parameters. For Kovač et al. (2014) we apply the offset between the color-based and SFR-based galaxy classification in order to plot the $f_{\rm Q}$ estimates. We also plot the $f_{\rm Q}(\mathcal{M}_{\rm fid} = 10^{10.5}\mathcal{M}_\odot)$ evolution of our sample with varying environment cut-offs specified on the top right. As in Figure 3.5 the width of the $f_{\rm Q}(\mathcal{M}_{\rm fid} = 10^{10.5}\mathcal{M}_\odot)$ evolution represent the uncertainty in the best-fit parameters of Equation 3.7. 143

4.1 The stellar mass function (SMF) that we use in our subhalo abundance matching (SHAM) prescription to construct galaxy catalogs from the Wetzel et al. (2013) `TreePM` simulation (§4.4). For our fiducial SMF (solid), we use the Li & White (2009) SMF at $z = 0.05$ and interpolate between the Li & White (2009) SMF and the Marchesini et al. (2009) $z = 1.6$ SMF for $z > 0.05$. To illustrate the evolution, we plot the SMF at $z = 0.05, 0.5$, and 0.9. We also plot a SMF parameterization using an "extreme" model of SMF evolution (dashed-dotted), in which the amplitude of the SMF at $z = 1.2$ is half the amplitude of the fiducial SMF. We later use this extreme model to ensure that the results in this work remain robust over different degrees of SMF evolution at $z > 0.05$. 160

4.2 The quiescent fraction of central galaxies, $f_{\mathrm{Q}}^{\mathrm{cen}}$, at $z < 1$ in different stellar mass bins. We compare our parameterzation of $f_{\mathrm{Q}}^{\mathrm{cen}}$ (Eq. 4.2) using the best-fit parameter values listed in Table 4.1 (dashed) to the $f_{\mathrm{Q}}^{\mathrm{cen}}$ measurements from Tinker et al. (2013) (scatter). For our parameterization, we fit $f_{\mathrm{Q}}^{\mathrm{cen}}$ at $z = 0.$ using central galaxies of the SDSS DR7 group catalog and fit $\alpha(\mathcal{M}_*)$, which dictates the redshift dependence, from the redshift evolution of the Tinker et al. (2013) measurements. 161

4.3 Schematic diagram that illustrate the star formation evolution of central galaxies in our model (§4.5.3). We plot SFR as a function of $t_{\rm cosmic}$ for a star forming galaxy (blue dashed), a star-forming galaxy that quenches at $t_{\rm Q,start} = 9$ Gyr (green) and mark the general region of quiescent galaxies (orange). Central galaxies while they are star-forming have SFRs that evolve with the SFMS, which decreases with cosmic time. When star-forming central galaxies quench, their SFR decreases exponentially with $t_{\rm cosmic}$. The quenching timescale, $\tau_{\rm Q}^{\rm cen}$, we constrain in our analysis dictates how rapidly these galaxies quench based on Eq. 4.8. For comparison we also plot the SFR evolution of a satellite with the same mass using the Wetzel et al. (2013) quenching timescales (red dashed). 167

4.4 **Top**: The evolution of the SSFR distribution in our model (§4.5.2) for a reasonable set of parameter values. The model evolves the SFR of central galaxies from $z \sim 1$ (light) to 0.05 (dark) while reproducing the observed SFMS and quiescent fraction evolutions. The shift in the star forming peak of the SSFR distribution from $z = 1$, reflects the overall decline in SFR of the SFMS over time. The quiescent fraction evolution is reflected in the growth of the quiescent peak accompanied by the decline of the star-forming peak. **Bottom**: Comparison of the SSFR distribution at $z = 0.05$ using a relatively shorter (dashed) and longer (dotted) quenching timescale than the above panel (solid). The quenching timescale (parameterized by A_τ and δ_τ), dictates how long quenching central galaxies spend in between the peaks. This is ultimately reflected in the height of the green valley. For longer quenching timescales, the height of the SSFR distribution green valley will higher. For shorter quenching timescales, it will lower. 176

4.5 We present the constraints we obtain for our model parameters using ABC-PMC. The diagonal panels plot posterior distributions of each of our model parameters, while the off-diagonal panels plot the degeneracies of parameter pairs. For each of the posterior distributions, we mark the 68% confidence interval (vertical dashed lines). We also mark the median of the posterior distributions in all the panels in black. 177

4.6 Comparison between the SSFR distribution calculated using the median of the ABC posterior distribution as the set of model parameters (orange) and the SSFR distribution of the SDSS DR7 central galaxies (black dash). The SSFR distribution from the median of the ABC posterior show good overall agreement. The distributions are especially consistent in the transition (green valley) regions, which are dictated by the quenching timescale parameters. . 178

4.7 Quenching timescale, $\tau_\mathrm{Q}^\mathrm{cen}$, of central galaxies (red) as a function of stellar mass. We plot $\tau_\mathrm{Q}^\mathrm{cen}$ of the median parameter values of the ABC posterior distributions (red points) along with $\tau_\mathrm{Q}^\mathrm{cen}$ drawn from the final iteration ABC parameter pool (faint red lines). For comparison, we also plot the satellite quenching timescale of Wetzel et al. (2014) (black dashed). The constraints we get for quenching timescale of central galaxies reveal that central galaxies have a significantly longer quenching timescale than satellite galaxies. 179

4.8 The SSFR distribution generated from the median values of the parameter constraints obtained from our analysis using Wetzel et al. (2013) satellite quenching timescale as the quenching timescale of our central galaxies (red) in four stellar mass bins. In each panel, we reproduce the quiescent fraction of the SDSS DR7 central galaxies; however, comparison to the SSFR distribution of the SDSS DR7 centrals (black dash) find significant discrepancies in each of the bins. The SSFR distribution using satellite quenching timescale have much shallower green valley regions as a result of galaxies quenching much faster with satellite quenching timescale. *This disagreement of model predictions for satellites applied to observations of centrals clearly demonstrates that centrals require longer quenching timescales than satellites.* 180

4.9 Central galaxy quenching timescales ($\tau_{\rm Q}^{\rm cen}$) derived from using SMF prescriptions with no SMF evolution (green) and with extreme SMF evolution (red) in our analysis. For comparison we include the satellite galaxy quenching timescale from Wetzel et al. (2013) and $\tau_{\rm Q}^{\rm cen}$ we obtain using our fiducial SMF prescription. Even extreme choices for the SMF evolution is insufficient to account for the significant difference between the central and satellite quenching timescales. The different SMF evolution mainly impacts the mass dependence, not the amplitude of $\tau_{\rm Q}^{\rm cen}$. 181

4.10 SSFR distributions of the central galaxies versus the satellite galaxies in the SDSS DR7 Group Catalog with stellar mass between $10^{10.1}$ and $10^{10.5}\mathrm{M}_{\odot}$. Both SSFR distributions have similar green valley heights (green shaded region). Since central galaxies have significantly longer quenching timescales, satellite galaxies have a higher rate of transitioning from star-forming to quenching than central galaxies. 190

4.11 Comparison of the central galaxy quenching migration time estimate we infer, ($t_{\rm mig}^{\rm cen}$; orange) with quenching time estimates for gas depletion absent accretion (strangulation) and morphological quenching. The width represents the 68 % confidence region propagated from the posterior distributions of the $\tau_{\rm Q}^{\rm cen}$ parameters. For strangulation, we include the gas depletion time at $z = 0.2$ derived from the star formation efficiency estimates in Popping et al. (2015) (blue dash-dotted). The surrounding blue shaded region plots the range of gas depletion times at $z = 0.15$ (longer) to 0.25 (shorter). We also include the quenching migration time inferred from the Peng et al. (2015) gas regulation model (dashed). For morphological quenching we plot the quenching times taken from the star formation histories of the simulated galaxy in Martig et al. (2009) (star). We also include the quenching times of the Milky Way in Haywood et al. (2016) (triangle). The quenching timescale of strangulation exhibit a similar stellar mass dependence and is generally consistent with our central quenching timescales. Although its feasibility for a wider galaxy population is unexplored, the quenching timescale from morphological quenching is in good agreement with our timescale. 191

12 Comparison of the SSFR distribution calculated using SSFRs from the SDSS DR7 group catalog (black dashed) versus Moustakas et al. (2013) (orange) for galaxies that are in both the SDSS DR7 group catalog and the Moustakas et al. (2013) sample $z \sim 0.1$ bin: $P(\mathrm{SSFR}^{\mathrm{group}})$ versus $P(\mathrm{SSFR}^{M2013})$. Galaxies are binned based on the group catalog $\mathcal{M}_*$ for both distributions so that the same galaxies are examined in each bin. We impose SSFR bounds on $P(\mathrm{SSFR}^{M2013})$ for low SSFRs to reproduce the $P(\mathrm{SSFR}^{\mathrm{group}})$ quiescent peak (§ 4.3). We note that the M2013 sample does not contain a large number of galaxies within the group catalog's z range at higher mass bins due its bright magnitude limit. We find good overall agremeent between $P(\mathrm{SSFR}^{\mathrm{group}})$ and $P(\mathrm{SSFR}^{M2013})$. Furthermore, they have consistent green valley heights, which is the main feature of $P(SSFR)$ critical for constraining the central quenching timescale. 204

13 The $(NUV - R) - (R - J)$ color-color relation for the SDSS DR7 group catalog (black) calculated from photometry compiled from GALEX, SDSS, and 2MASS (§ 4.8). Galaxies classified as quiescent using the $\mathrm{SFR} - \mathcal{M}_*$ cut are highlighted (orange). Furthermore, we plot the color-color cuts from Bundy et al. (2010) that describe the classification of star-forming/quiescent galaxies in Tinker et al. (2013). We note that the quiescent galaxies classified using the $\mathrm{SFR} - \mathcal{M}_*$ cut are generally consistent with the Bundy et al. (2010) color-color cuts. 207

14 Comparison of the SDSS DR7 group catalog $f_{\rm Q}(\mathcal{M}_*)$ measured using the SFR $-$ $\mathcal{M}_*$ versus the $(NUV - R) - (R - J)$ color-color classifications. The $f_{\rm Q}$s measured using the two different classification methods are consistent with each other. This consistency illustrates that the $f_{\rm Q}^{\rm cen}(\mathcal{M}_*, z)$ parameterization derived from measurements of SDSS DR 7 group catalog and Tinker et al. (2013) does *not* affect the results of this work. 208

List of Tables

1.1 d_{LOS} Distribution Best-fit Parameters 44

2.1 **Prior Specifications**: The prior probability distribution and its range for each of the Zheng et al. (2007a) HOD parameters. All mass parameters are in unit of $h^{-1}M_{\odot}$ 95

3.1 Galaxy Subsamples 121

3.2 Best Fit Parameters for $f_{\mathrm{Q}}(\mathcal{M}_{*})$ Fit 140

4.1 Parameterizations in the Central Galaxy SFH Model with Fixed Parameters 175

4.2 Parameterizations in the Central Galaxy SFH Model with Free Parameters . 178

List of Appendices

Appendix A **201**

Appendix B **202**

Appendix C **205**

Introduction

Amidst the countless stars and galaxies we observe in the Universe lie undetected structures of dark matter, orders of magnitude larger than the luminous objects they engulf. These vast invisible structures began in the very early Universe, as quantum fluctuations in the aftermath of the Big Bang. During the subsequent period of inflation, these primordial fluctations were amplified by the accelerated expansion of the Universe and then propagated through gravitational instability for billions of years.

Despite constituting most of the matter in the Universe, dark matter has yet to be directly observed. In fact, it can only be studied through its gravitational interactions with luminous baryons, the matter of stars, galaxies, and celestial objects that emit light. In a way, the galaxies we observe in the cosmic volumes probed by our telescopes act as illuminated beacons tracing the vast dark matter terrains of the Universe.

Over the past decade, spectroscopic redshift surveys like the Sloan Digital Sky Survey III Baryon Oscillation Spectroscopic Survey (BOSS; Anderson et al. 2012; Dawson et al. 2013a) have exploited these galactic beacons to map out cosmic structures of the Universe. Precise measurements of distance and growth of large-scale structure (LSS) from these surveys, provide tests of cosmological models that describe the content, geometry and history of the Universe. The next leap in galaxy surveys will continue to expand the cosmic volumes probed by galaxies. These observations have the potential to constrain cosmological parameters with unprecedented precision. In the following sections, I briefly introduce how observations from galaxy redshift surveys can be used to test cosmological models, General Relativity, and particle physics beyond the Standard Model.

0.1 Large Scale Structure in ΛCDM

From the early Universe, primordial quantum fluctuations grow into the large-scale structures of the Universe we observe today through gravitational instability over different epochs of cosmic history. In this section, I briefly describe the simplified (*linear*) theory of this evolution and explain core concepts of LSS cosmology using galaxies. Let us begin by defining the matter overdensity field (or density fluctuation) at comoving position $\boldsymbol{r}$:

$$\delta(\boldsymbol{r}) = \frac{\rho(\boldsymbol{r}) - \bar{\rho}}{\bar{\rho}}, \tag{1}$$

where $\rho(\boldsymbol{r})$ is the density field and $\bar{\rho}$ is the mean density. Then, in Fourier space the density fluctuation becomes

$$\delta(\boldsymbol{k}) = \int \frac{\mathrm{d}^3\boldsymbol{r}}{(2\pi)^3}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\, \delta(\boldsymbol{r}), \tag{2}$$

the Fourier transform of $\delta(\boldsymbol{r})$. For describing the evolution of the overdensity field, Fourier space is often favored in the literature. The information in the overdensity field is often quantified using its N-point statistics (Peebles, 1980a; Bernardeau et al., 2002; Dodelson, 2003). In fact, the two-point statistic is one of the most commonly used tool in large scale structure studies. This two-point statistic, which is also referred to as the correlation function, is defined as

$$\xi(\boldsymbol{r}) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r}) \rangle \tag{3}$$

and in Fourier space as

$$\langle \delta(\boldsymbol{k})\delta(\boldsymbol{k'})\rangle = (2\pi)^3 P(\boldsymbol{k})\, \delta^D(\boldsymbol{k}+\boldsymbol{k'}). \tag{4}$$

δ^D is the Dirac delta function and $P(\boldsymbol{k})$ is the two-point statistic in Fourier space – the *power spectrum*. $P(\boldsymbol{k})$ is the Fourier transform of $\xi(\boldsymbol{r})$ and, in principle, they contain the same information. In practice, however, analyzing $\xi(\boldsymbol{r})$ and $P(\boldsymbol{k})$ carry different caveats (Feldman et al., 1994). Throughout this dissertation, I will mainly focus on the power spectrum.

Now in order to determine the evolution of the matter overdensity field (on sub-horizon scales) consider pressureless dark matter, which consistutes most of the matter in the Universe. From the continuity, Euler, and Poisson equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho\, \boldsymbol{u} = 0 \tag{5}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\cdot\boldsymbol{u} - \nabla\Phi = 0 \tag{6}$$

$$\nabla^2\Phi - 4\pi G\rho = 0 \tag{7}$$

its equation of motion can be derived

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} - 4\pi G\bar{\rho}\,\delta = 0. \tag{8}$$

$\boldsymbol{u}$ is the velocity field, Φ is the gravitational potential, and a is the scale factor. For a detailed derivation I refer readers to Peebles (1980a) and Dodelson (2003). The solution for this second order differential equation can be written as

$$\delta(\boldsymbol{r},t) = D^{(+)}(t)A(\boldsymbol{r}) + D^{(-)}(t)B(\boldsymbol{r}). \tag{9}$$

The density flucation has two components: a growing mode $D^{(+)}$ and a decaying mode $D^{(-)}$. The decaying mode, as its name suggests, decreases with time and its contribution becomes negligible in the late Universe leaving only the growing mode. To quantify the evolution of the growing mode $D^{(+)}$, one commonly used quantity is the "growth rate of structure":

$$f = \frac{d \ln D^{(+)}}{d \ln a}. \tag{10}$$

This growth rate of structure is a key quantity in LSS cosmology for testing different cosmological models and theories of gravity. f will be discussed further in Section 0.2.

In addition to their gravitation evolution, the density fluctuations evolve through different epochs in cosmic history: inflation, radiation-dominated era, matter-radiation equality, and matter-dominated era. Each of these periods leave an imprint on the evolution of δ. In Fig. 1, I mark the eras in the early Universe and plot how the physical scale of the Universe, represented by the Hubble radius, evolves with the scale factor a.

During inflation, the Hubble radius remains constant. Afterwards the Universe becomes radiation dominated. Based on the Friedmann equations the Hubble radius during the radiation dominated era is approximately $\propto a^2$. After a period when radiation and matter have comparable energy densities, the Universe becomes matter dominated where the Hubble radius is approximately $\propto a^{3/2}$. Meanwhile, the physical scale of perturbations is $\lambda_{phys} = \lambda_{comov}\, a(t)$ and thus $\propto a(t)$. As Fig. 1 schematically illustrates, perturbations exit the Hubble radius during inflation then reenter the Hubble radius later on. Depending on the physical scale of the perturbation, it enters either during the radiation-dominated era (smaller scale) or matter-dominated era (larger scale).

The physical scale of perturbations that enter the horizon at the time of matter-radiation equality, where $a(t) = a_{eq}$, is $\lambda_{eq} \sim 500\ h^{-1}$Mpc. Then perturbations that enter before the matter-radiation equality have physical scales $\lambda_{phys} < \lambda_{eq}$ and since they enter during the radiation dominated era, these smaller scale perturbations are effectively frozen and hence their growth is suppressed. On the other hand, the larger scale perturbations with $\lambda_{phys} > \lambda_{eq}$ enter after matter-radiation equality during the matter dominated epoch. These perturbations do *not* experience the suppression of growth of the radiation dominated era. The net effect on the overdensity as it goes through these epochs is the suppression of growth on scales smaller than λ_{eq}, or k_{eq} in Fourier space, by a factor of $\sim k^4$. In practice, this scale dependent evolution of the density fluctuation is quantified through the "transfer function" $T(k)$ (Eisenstein & Hu, 1998, 1999).

The density fluctuations after inflation can be summarized by the power spectrum:

$$P_{\rm inf}(k) \propto k^{n_s} \tag{11}$$

where n_s, the spectral tilt of the primordial power spectrum, is measured to be ~ 1 (Harrison, 1970; Peebles & Yu, 1970; Zeldovich, 1972; Komatsu et al., 2011). Then, the power spectrum of the density fluctuation in the late Universe can be expressed as

$$P(k) \propto k^{n_s}\ T^2(k)\ D^2(k). \tag{12}$$

where $D(k) \equiv D^{(+)}$ is the growth function from earlier this section. Through the cosmological models and parameters, which predict $T(k)$ and $D(k)$, we predict the power spectrum of the density fluctuation. Then these predictions can then be compared to measurements made from observations in order to produce constraints on cosmological parameters, better understand dark energy, and test theories of gravity.

Unfortunately, most of the matter in the Universe is in the form of dark matter and does not interact with radiation, so observers cannot measure the spatial/clustering statistics

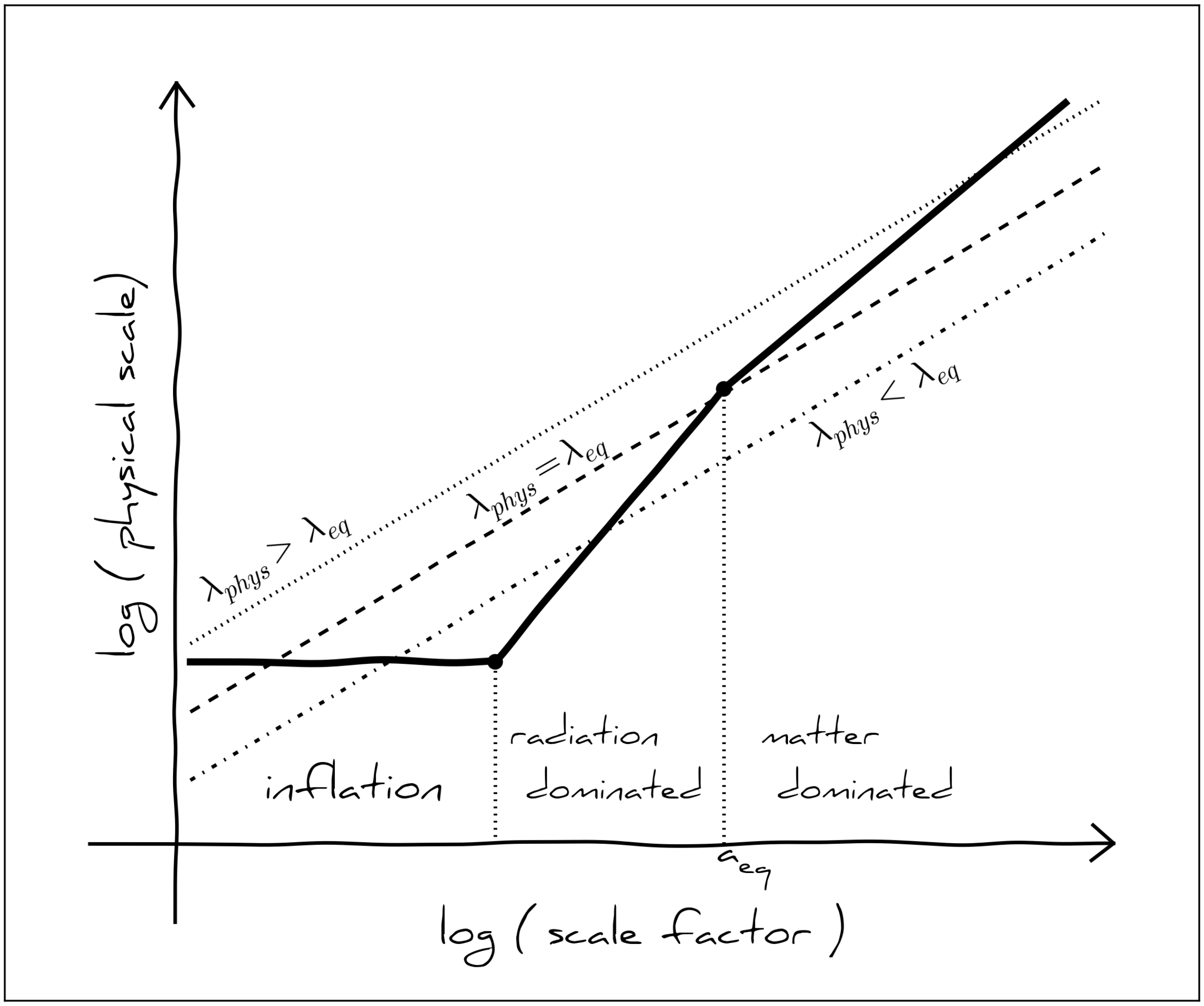
log (physical scale)
$\lambda_{phys} > \lambda_{eq}$
$\lambda_{phys} = \lambda_{eq}$
$\lambda_{phys} < \lambda_{eq}$
inflation
radiation dominated
matter dominated
a_{eq}
log (scale factor)

Figure 1: Schematic diagram that illustrates the evolution of the density fluctuations in the early Universe through inflation, radiation-dominated epoch, matter-radiation equality, and matter-dominated epoch. The evolution of the Hubble radius (solid line) remains flat during inflation (flat), scales by $\propto a^2$ during radiation domination, and $\propto a^{3/2}$ during matter domination. a_{eq} marks matter radiation equality. The physical lengths of three constant comoving scales are marked by dashed, dotted, and dot-dashed lines. The dashed line represents physical lengths of perturbations that enter the horizon during matter-radiation equality $\lambda_{phys} = \lambda_{eq}$. The dot-dashed line mark perturbations that enter the horizon during radiation domination with $\lambda_{phys} < \lambda_{eq}$. The dotted line mark perturbations that enter the horizon during matter domination with $\lambda_{phys} > \lambda_{eq}$. As described in the text, the growth of perturbations with $\lambda_{phys} < \lambda_{eq}$ are suppressed because they enter the horizon during the radiation dominated era. The evolution of the density perturbation through these epochs are quantified through the transfer function $T(k)$.

of dark matter directly. Instead, we measure the clustering of galaxies or quasars, which trace the underlying matter distribution. The smoothed galaxy/quasar density field can be approximated by a local function of the matter density field

$$\delta_g(\boldsymbol{r}) = f(\delta(\boldsymbol{r})). \tag{13}$$

$f(\delta(\boldsymbol{r}))$ can then be expanded Taylor series (Fry & Gaztanaga, 1993):

$$\delta_g(\mathbf{r}) = \sum_{k=0}^{\infty} \frac{b_k}{k!}\delta^k \tag{14}$$

where b_0 is chosen so that $\langle \delta_g \rangle = 0$ and b_1 is referred to as the linear bias factor. To linear order,

$$P_g(k) = b_1^2 P(k). \tag{15}$$

The primary galaxy subpopulation used in LSS studies so far are luminous red galaxies (Eisenstein et al., 2001; Dawson et al., 2013a). These galaxies have $b_1 > 1$, which makes them *biased* tracers of the matter distribution (Zehavi et al., 2005; Sheldon et al., 2009; Gaztañaga et al., 2009; Zhai et al., 2016). Luminous galaxies reside in larger potential wells. The peaks of the density fluctuation have stronger clustering properties than the overall overdensity field (Manera et al., 2010).

Based on the derivation of this section, once we have the spatial distribution of galaxies or quasars, we can derive the clustering of the matter distribution and then infer cosmological constraints. In practice, however, a number of factors complicate this procedure. One major complication is redshift-space distortions, which will be discussed in the next section.

0.2 Redshift-Space Distortions

Spectroscopic redshifts surveys, such as 2dF Galaxy Redshift Survey (Colless, 1999), Sloan Digital Sky Survey (SDSS York et al., 2000), and BOSS, have mapped out millions of distant galaxies. Current surveys such as Extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2015), and future surveys such as the Dark Energy Survey Instrument (DESI; Schlegel et al. 2011; Morales et al. 2012; Makarem et al. 2014) and the Subaru Prime Focus Spectrograph (PFS; Takada et al. 2014), will continue to map out millions more. These surveys dominate LSS studies and have/will been critical for inferring precise cosmological constraints. As their name suggest, however, these *redshift* surveys do not directly measure the actual position of galaxies. Instead they measure the angular positions (right ascension and declination) and redshifts of galaxies.

These redshifts are a combined measurement of the recession velocities due to the expansion of the Universe and the peculiar velocities of the galaxies:

$$z_{\rm obv} = z_{\rm true} + \frac{v_{\rm pec}}{c}. \tag{16}$$

The galaxy comoving positions derived from the angular positions and redshifts are in *redshift-space* and "distorted" compared to real-space comoving positions by

$$\boldsymbol{s} = \boldsymbol{x} + \frac{\boldsymbol{v}_{\rm pec} \cdot \hat{n}}{H_0} \tag{17}$$

where $\hat{n}$ is the unit vector along the line-of-sight. Thankfully, all hope is not lost.

The peculiar velocities of galaxies are directly related to the total matter distribution, since galaxies can be thought of as test particles in a gravitational field. Using this relation, Kaiser (1984) derived an approximation for the distortion caused by the coherent infall of galaxies onto overdense regions in redshift space. This redshift-space distortion (RSD), often referred to as the Kaiser effect, causes observations of overdense regions to appear squashed along the line of sight in redshift-space. Galaxies around an overdense region that are closest to the observer on Earth are moving towards the center of the overdense region and away from the observer. So in they appear farther away than their true position. Galaxies on the other side are moving towards both the overdense region and the observer, so they appear closer to us in redshift-space.

The relation between the overdensity field in redshift-space can be derived from the continuity equation and the distant observer approximation,

$$\delta^{(s)}(\boldsymbol{k}) = (1 + f\mu^2)\delta(\boldsymbol{k}). \tag{18}$$

f is the growth rate of structure (Eq. 10) and $\mu = \boldsymbol{k} \cdot \hat{n}/k$, cosine of the angle between $\boldsymbol{k}$ and the line-of-sight. In corporating the Kaiser effect into the galaxy bias model from Section 0.1, the galaxy/quasar overdensity field in redshift-space becomes

$$\delta_g^{(s)}(\boldsymbol{k}) = (b_1 + f\mu^2)\delta(\boldsymbol{k}). \tag{19}$$

The redshift-space power spectrum of the galaxy overdensity field can then be written as

$$P_g^{(s)}(k, \mu) = (b_1 + f\mu^2)^2 P(\boldsymbol{k}). \tag{20}$$

On large scales and with small overdensities, the effect of redshift-space distoritons is well described by the Kaiser effect. On small scales with large overdensities things get a little more complicated.

The random peculiar velocities of galaxies in gravitationally bound structures such as clusters cause their position in redshift-space to be smeared out to larger scales along the line-of-sight. This effect can easily be identified by eye in galaxy redshift maps where the elongations of the galaxy positions along the line-of-sight resemble fingers pointing towards the observer. Aptly this redshift-space distortion is referred to as the "fingers-of-god". Its impact on the power spectrum, is empirically modeled and typical quantified using an overall exponential factor (Jackson, 1972; Scoccimarro, 2004; Taruya et al., 2010; Beutler et al., 2016). Including both RSDs from the Kaiser effect and the fingers-of-god, the redshift-space power spectrum is then

$$P_g^{(s)}(k,\mu) \approx e^{-f^2\sigma_v^2\mu^2k^2}(b_1 + f\mu^2)^2 P(k) \tag{21}$$

where σ_v is a paramter quantifying the strength of the effect and is usually left as a free parameter in analyses.

Eq. 21 reveals the f dependence in RSDs. RSD analyses in LSS studies exploit this dependence by measuring the impact of RSDs on the power spectrum to constrain f. Consider the Legendre expansion of $P_g^{(s)}(k,\mu)$,

$$P_g^{(s)}(k,\mu) = \sum_{\ell=0,2,4...} \mathcal{L}_\ell(\mu) P_g^\ell(k). \tag{22}$$

Each of power spectrum "multipole" of this expansion is then

$$P_g^\ell(k) = \frac{2\ell+1}{2}\int\limits_{-1}^{1} \mathrm{d}\mu\; P_g^{(s)}(k,\mu)\;\mathcal{L}_\ell(\mu). \tag{23}$$

The RSD factor in the power spectrum multipoles for $\ell = 0$ (monopole) and 2 (quadrupole) are

$$P_g^0(k) = (b_1^2 + \frac{2}{3}fb_1 + \frac{1}{5}f^2)P(k) \quad (24)$$

$$P_g^2(k) = (\frac{4}{3}fb_1 + \frac{4}{7}f^2)P(k). \quad (25)$$

For simplicity, we neglect the fingers-of-god, which does not significantly impact larger scales. Taking the ratio of the quadrupole over the monopole,

$$\frac{P_g^2}{P_g^0} = \frac{\frac{4}{3}fb_1 + \frac{4}{7}f^2}{b_1^2 + \frac{2}{3}fb_1 + \frac{1}{5}f^2}, \quad (26)$$

we can in principle eliminate the dependence on scale and extract information on f. Of course in practice the simplified derivations of this section break down. Instead of the simple linear theory theoretical models I derived, models of $P_g^{(s)}$ are derived using perturbation theory and incorporate more sophisticated RSD and bias models (Bernardeau et al., 2002; Scoccimarro, 2004; Taruya et al., 2010; Nishimichi & Taruya, 2011; Taruya et al., 2013, 2014; Beutler et al., 2016). These models are then compared to the observed P_g multipoles from galaxy surveys in order to derive constraints on cosmological parmaeters such as f.

0.3 Weighting Neutrinos with Galaxies

Beyond inferring the growth rate of structure, which can be used to test GR and modified gravity scenarios, galaxy clustering also provides a unique window to probe fundamental physics beyond the standard model. In the derivations of Sections 0.1 and 0.2 we focused on how the dark matter density fluctuations of evolves. This is an excellent approximation because dark matter consistutes the majority of matter in the Universe. However, it negelects some of the more detailed imprints on LSS from other components of matter – *i.e.*

neutrinos, which oscillation and detection experiments have *very* convincingly (Nobel Prize in Physics 2015) confirmed is *not* massless (Hu & Eisenstein, 1998; Lesgourgues & Pastor, 2012; Lesgourgues et al., 2013; Lesgourgues & Pastor, 2014).

In the very early Universe, neutrinos are relativistic and coupled to the primordial plamsa. Later they decouple from the plasma, while they are still ultra-relativistic and redshift. At this point, they do not contribute to the energy density of matter but instead radiation. Eventually during matter domination era, neutrinos become non-relativistic and then contribute to the matter energy density acting as "warm/hot" dark matter. After decoupling from the primoridal plasma, neutrinos are effectively a collisionless fluid, where the individual particles free-stream with characteristic velocities defined by their thermal velocity. Earlier on when they are relativistic, their free-streaming scale is simply equal to the Hubble radius. Later when they are non-relativistic, their characteristic velocity is approximately

$$v_{\rm th} \approx 158(1+z)\left(\frac{1{\rm eV}}{m}\right) \quad {\rm km\ s}^{-1} \tag{27}$$

and the free-streaming scale can be derived in an analogous way as the Jean's length derivation:

$$\lambda_{\rm FS} = 2\pi\sqrt{\frac{2}{3}}\left(\frac{v_{\rm th}}{H}\right) \tag{28}$$

or

$$k_{\rm FS} = \frac{2\pi a}{\lambda_{\rm FS}} \approx 0.82\frac{\sqrt{\Omega_\Lambda + \Omega_m(1+z)^3}}{(1+z)^2}\left(\frac{m_\nu}{1\ {\rm eV}}\right). \tag{29}$$

where Ω_Λ and Ω_m are the current cosmological constant and matter density fractions, respectively.

Neutrinos leave two main imprints on LSS. In the early Universe they contribute to the radiation energy density but later, they contribute to the matter energy density. As described in Section 0.1, matter-radiation equality marks the turning point in suppression of growth of structure, quantified by $T(k)$. The transition of neutrinos from radiation to matter impacts a_{eq} and thus impacts $T(k)$ by shifting the turning point of the cold dark matter (CDM) only power spectrum. Even after becoming non-relativistic, neutrinos still do not contribute to the clustering of matter on scale smaller than $k_{\rm FS}$. The impact of this scale dependent suppression of clustering, can be analytically estimated for the matter power spectrum (Bird et al., 2012):

$$\frac{\Delta P}{P} = \frac{P^{f_\nu \neq 0} - P^{f_\nu = 0}}{P^{f_\nu = 0}} \approx -8 f_\nu \quad \text{for} \;\; k \gg k_{\rm FS}. \tag{30}$$

where f_ν is the ratio of the neutrino energy density over that of matter (Ω_ν/Ω_m).

The total mass of neutrinos (Σm_ν) dictates the strength of these imprints and can therefore be constrained by the shape of the power spectrum. The same tools used for analyzing RSDs and measuring the growth rate of structure can also be used to measure Σm_ν from observations of galaxy surveys (Hu et al., 1998; Costanzi et al., 2013; Villaescusa-Navarro, 2015; Cuesta et al., 2016a). Based on forecasts, the next galaxy surveys such as DESI[1] have the potential to infer the most stringent constraints on Σm_ν– $\sigma_{\sum m_\nu} \sim 0.03$ eV. Such constraints have the potential to distinguish between the normal or invereted neutrino mass hierarchy and reveal physics beyond the Standard Model.

0.4 Analyzing Galaxy Clustering

In the previous sections, I laid out the theoretical framework for LSS analysis using galaxy clustering. As I alluded earlier, the models and predictions of this theoretical framework can be compared to observations to derive constraints on parameters of interest. In this section

[1] DESI *Final Design Report* (FDR): http://desi.lbl.gov/tdr/

I describe the statistical framework for comparing the theoretical models to observations from galaxy surveys. The ultimate goal of galaxy clustering analyses is to derive probability distributions of the cosmological parameters (*e.g.* f, Σm_ν) given the data from observations. The standard approach to deriving this *posterior* probability distribution is using Bayesian parameter inference. Based on Bayes theorem, the posterior probability distribution can be expressed as

$$P(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{P(\boldsymbol{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{D})}. \tag{31}$$

$\boldsymbol{D}$ and $\boldsymbol{\theta}$ refer to observations and cosmological parameters, respectively. $P(\boldsymbol{D}|\boldsymbol{\theta})$, the probability distribution function for the observation $\boldsymbol{D}$ given model parameters $\boldsymbol{\theta}$, is the *likelihood function* ($\mathcal{L}$). $P(\boldsymbol{\theta})$ is the *prior* probability distribution function. Lastly, $P(\boldsymbol{D})$ is the "evidence", which for our purposes is just a normalization factor independent of $\boldsymbol{\theta}$. The equation is more commonly simplified as

$$\begin{aligned} P(\boldsymbol{\theta}|\boldsymbol{D}) &\propto P(\boldsymbol{D}|\boldsymbol{\theta})\; P(\boldsymbol{\theta}) && (32)\\ \text{posterior} &\propto \text{likelihood} \times \text{prior.} && (33) \end{aligned}$$

In the context of galaxy clustering analyses and LSS cosmology in general, the likelihood function is *typically* assumed to have Gaussian function form and calculated as

$$P(\boldsymbol{D}|\boldsymbol{\theta}) = \mathcal{L} = \frac{1}{(2\pi)^{N_d/2}\,\mathrm{det}\boldsymbol{C}^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{D} - F(\boldsymbol{\theta}))^T \boldsymbol{C}^{-1}(\boldsymbol{D} - F(\boldsymbol{\theta}))\right]. \tag{34}$$

$\boldsymbol{D}$ is data observed and measured from galaxy surveys with dimension N_d. $F(\boldsymbol{\theta})$ is the model prediction of the observable (*e.g.* $P_g^{(s)}$) generated from cosmological parameters $\boldsymbol{\theta}$, described in earlier sections. And $\boldsymbol{C}$ is the covariance matrix.

A number of different methods are used to estimate the covariance matrix. For instance, efforts to analytically estimate the covariance matrix from theory have been made in the past (Hamilton et al., 2006; Pope & Szapudi, 2008; de Putter et al., 2012). However, non-linear evolution, shot-noise, RSDs, and mapping between galaxies and matter complicate accurate estimations. Jack-knife resampling (Shao & Tu, 1995), a commonly used method in astronomy for estimating covariances directly from data have also been used. However, the method requires a number of arbitrary choices and cannot account for fluctuations on the scale of the survey (Norberg et al., 2009). Instead, the latest analyses estimate $\boldsymbol{C}$ from galaxy mock catalogs generated from N-body simulations. For accurate estimation, an order of ~ 1000 mock galaxy catalogs are required in the analysis (Scoccimarro & Sheth, 2002; **?**; Anderson et al., 2012; Manera et al., 2013; Rodríguez-Torres et al., 2015; Kitaura et al., 2016; Beutler et al., 2016) Developing fast and accurate galaxy mock catalogs for LSS analyses has now become a subfield of its own. As an added detail, in standard analyses, in order to account for biases in the $\boldsymbol{C}$ estimates, include a correction – the Hartlap factor – to the covariance matrix estimate (Hartlap et al., 2007).

From $\boldsymbol{D}$, $F(\boldsymbol{\theta})$, and $\boldsymbol{C}$ we can evalulate an estimate of the likelihood function. From the likelihood, since the prior probability distribution is chosen *a priori*, the posterior probability distribution functions of the cosmological parameters is essentially already evaluated. In practice, the posterior distribution is not evaluated at all points in parameter space, but rather sampled using a sampler such as a Markov Chain Monte Carlo sampler (*e.g.* `emcee` Foreman-Mackey et al., 2013).

From the galaxy clustering analysis described in this chapter, the latest galaxy surveys have produced some remarkable constraints on cosmological parameters. From the SDSS and BOSS surveys, measurements of the power spectrum multipoles along with analogous configure-space analyses have yielded a number of constraints on $f\sigma_8$ (Reid et al., 2012; Oka et al., 2014; Beutler et al., 2014b; Alam et al., 2015, 2016; Beutler et al., 2016), where σ_8

is the the rms linear fluctuation in density perturbations on scales of 8 h^{-1}Mpc. Similar to multipoles, power spectrum wedges have also been used, in both Fourier and configuration-spaces, to infer $f\sigma_8$ constraints (Sánchez et al., 2013; Sanchez et al., 2016; Grieb et al., 2016). These $f\sigma_8$ constraints can then be compared to cosmological predictions from Cosmic Microwave Background (CMB) experiments such as the Wilkinson Microwave Anisotropy Probe (Hinshaw et al., 2013) and *Planck* (Planck Collaboration et al., 2014a) to test ΛCDM cosmology and General Relativity. The constraints from BOSS are generally consistent with ΛCDM and GR over $0.2 < z < 0.75$. For instance, Beutler et al. (2016) derives $f\sigma_8 = 0.482 \pm 0.053, 0.455 \pm 0.050$, and 0.410 ± 0.042 from BOSS for effective redshift $z_{\text{eff}} = 0.38, 0.51$, and 0.61. $f\sigma_8$ constraints from galaxy power spectrum analyses have also been combined with CMB data to constrain Σm_ν (Zhao et al., 2013; Beutler et al., 2014a; Gil-Marín et al., 2015). Beutler et al. (2014a), from combining constraints from galaxy power spectrum analyses with *Planck* CMB results, derives the upper bound $\Sigma m_\nu < 0.51$ eV.

Ongoing and future surveys, such as eBOSS, PFS, and DESI, will continue to collect many more million redshifts and expand the probed cosmic volume by an order of magnitude. These observations have the potential to produce cosmological parameter constraints with unprecedented statistical precision. The main challenges for realizing their full potential are *methodological*.

So far I have focused on LSS analyses using only the galaxy power spectrum – the two-point statistic of the density fluctuations. Analyses restricted to just the two-point statistic, however, face a number of limitations. The constraints on the growth rate of structure, listed above, have all constrained $f\sigma_8$ rather than f alone. The degeneracy between f and σ_8 cannot be broken with $P(k)$ alone. Furthermore, the $P(k)$ multipoles in Eq. 26 illustrate that $P(k)$ analyses also suffer from the degeneracy between f and bias parameters.

The *bispectrum* $B(k_1, k_2, k_3)$, the three-point statistic of density fluctuations, can be used

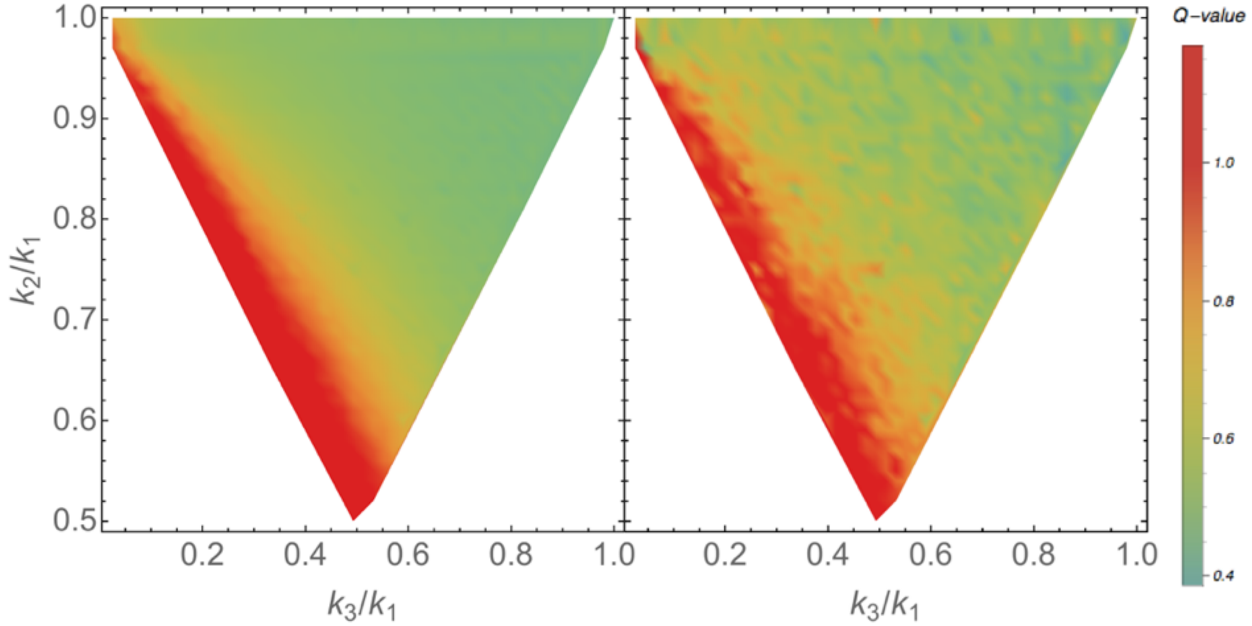

Figure 2: Amplitude of the reduced galaxy bispectrum $Q(k_1, k_2, k_3)$ plotted as a function of ratios k_2/k_1 and k_3/k_1, which describe the triangle configurations. The $Q(k_1, k_2, k_3)$ in the left panel is calculated from a perturbation theory model while the right panel presents $Q(k_1, k_2, k_3)$ of BOSS Data Release 12 CMASS galaxy sample using the Scoccimarro (2015) estimator.

to break the degeneracies among f, σ_8, and bias parameters (Scoccimarro et al., 1998; Verde et al., 1998; Scoccimarro, 2000, see Bernardeau et al. 2002 for a review). The dependence on triangle configuration in $B(k_1, k_2, k_3)$ disentangles contributions from gravitational instability versus non-linear biasing of galaxies. Without going into any further detail, in Figure2 I present the reduced galaxy bispectrum $Q(k_1, k_2, k_3) = B(k_1, k_2, k_3)/(P(k_1)P(k_2) + P(k_2)P(k_3) + P(k_1)P(k_3))$ measurement for the BOSS Data Release 12 CMASS galaxy sample (right) and a perturbation theory model (left). The BOSS $Q(k_1, k_2, k_3)$ is measured using the Scoccimarro (2015) estimator. $P(k)$ and $B(k_1, k_2, k_3)$ measurements from galaxy surveys can be jointly analyzed in order to derive constraints explicitly on f.

All LSS analyses suffer from observational systematic effects. For fiber-fed multi-object spectroscopic surveys (*e.g.* SDSS, BOSS, eBOSS, DESI, and PFS) these effects include variations in target selection relatd to stellar density, image depth, seeing, and other factors (Ross et al., 2012; Anderson et al., 2012). If not accounted for fiber collisions, for instance, prevent surveys from collecting a significant fraction redshifts due to physical constraints on the focal plane. As I detail in Chapter 1, their impact on $P(k)$ goes well beyond their angular scale and restricts analysis on small scales, which have higher signal-to-noise. In addition to diminishing the statistical power of galaxy redshift surveys, fiber collisions can also bias constraints on cosmological parameters. Many efforts have been made to tackle these challenges from observational systematics (Ross et al., 2012; Guo et al., 2012a, and Chapter 1).

In Eq. 34, the likelihood function assumes a Gaussian functional form – a standard assumption in LSS analyses. However, in detail, this assumption cannot be correct due to nonlinear gravitational evolution and biasing (Mo & White, 1996; Somerville et al., 2001; Casas-Miranda et al., 2002; Bernardeau et al., 2002). The likelihood also relies on the estimated covariance matrix to capture the sample variance of the data. Besides the labor and computational costs required to make them, simulated mock catalogs used for covariance matrix estimation are inaccurate on small scales (see Heitmann et al. 2008; Chuang et al. 2015a and references therein). Furthermore, using covariance matrix estimates rather than the "true" covariance matrix (Sellentin & Heavens, 2016) along with systematics impact the likelihood in ways difficult to model. Fortunately, evaluating the explicit likelihood is *not* necessary for inferring cosmological parameters. Likelihood-free inference techniques such as Approximate Bayesian Computation (ABC) relax these restrictions and make inference possible without making any assumptions on the likelihood. In Chapter 2 I combine ABC with a Population Monte Carlo sampler and apply it in the context of LSS.

As described earlier, galaxies are biased tracers of the underlying matter distribution. For more than a decade, halo occupation modeling has been a popular framework for connecting galaxies to the dark matter structures underneath in galaxy formation and cosmology studies (Yang et al., 2003; Tinker et al., 2005a; van den Bosch et al., 2007; Zheng et al., 2007b; Conroy & Wechsler, 2009a; Guo et al., 2011; Leauthaud et al., 2012a; Tinker et al., 2013; Zu & Mandelbaum, 2015). The standard halo occupation model assumes that galaxies reside in dark mater halos and their occupation is a function of only the mass of the halo. However, the clustering of dark matter halos depend on properties beyond their masses, such as their assembly history. If this effect, coined *halo assembly bias*, propagates to galaxies, it will induce *galaxy assembly bias* on standard halo occupaiton model and significantly impact galaxy clustering analyses (Hearin et al., 2016b; Zentner et al., 2016; Vakili & Hahn, 2016). Therefore, better understanding of the galaxy-halo connection is essential for LSS analyses.

Beyond their utility as tracers for cosmology, galaxies also pose fundamental questions regarding how the early homogenous Universe became the heterogenous one today. Observations have now firmly established a global view of galaxy properties out to $z \sim 1$ (*e.g.* Bell et al., 2004; Bundy et al., 2006a; Cooper et al., 2007; Cassata et al., 2008; Blanton & Moustakas, 2009a; Whitaker et al., 2012; Moustakas et al., 2013, Chapter 3). Galaxies roughly fall into two categories: star-forming galaxies and quiescent ones with little star formation. The star-forming population undergoes significant decline in star formation rate (SFR) over cosmic time and significant fractions of them also rapidly "quench" their star formation and become quiescent. The underlying drivers of this evolution, however, are not directly revealed by observations. Cosmology, which precisely predicts the dark matter evolution, provides a framework for answering *specific* and *tractable* questions in galaxy evolution. In ΛCDM, structures form "hierarchically" — smaller ones form earlier and subsequently merge to form larger ones. The galaxy population can be positioned in this framework with halo occupation in order to constrain key elements of their evolution and better understand the galaxy-halo connection (Wetzel et al., 2013, 2014; Tinker et al., 2016, 2017). In Chapter 4, I use this approach to measure the timescale of star-formation quenching in central galaxies.

In this dissertation, I tackle key methodological challenges in LSS analyses with galaxy clustering by developing methods to robustly treat systematics (Chapter 1), introducing innovative approaches to inference in LSS studies (Chapter 2), and improving our understanding of the galaxy-halo connection (Chapters 3 and 4). Each Chapter contributes to unlocking the full potential of current and future galaxy redshift surveys and will be critical for testing cosmological models and General Relativity and constraining the total neutrino mass.

Chapters 1 and 3 have both been refereed and published in the astronomical literature. Chapters 2 and 4 have both been refereed and accepted to the *Monthly Notices of the Royal Astronomical Society* and *The Astrophysical Journal*, respectively. All of these Chapters were co-authored with collaborators but the majority of the work and writing in each Chapter is mine. Below, I describe my contributions to each Chapter:

1. For Chapter 1, I developed the idea for the project in collaboration with Roman Scoccimarro and Michael Blanton. I implemented the project with contributions from

Roman Scoccimarro. The project utilized simulation data from Jeremy Tinker and Sergio Rodríguez-Torres. I wrote the paper with additions from Roman Scoccimarro and edits by Michael Blanton.

2. For Chapter 2, I developed the idea for the project in collaboration with Mohammadjavad Vakili, Andrew Hearin, and David Hogg. I implemented the project with Mohammadjavad Vakili and contributions from Andrew Hearin and Kilian Walsh. The project utilized software written by Andrew Hearin and Duncan Campell. I wrote the paper together with Mohammadjavad Vakili with additions from Andrew Hearin, David Hogg, and Kilian Walsh.

3. For Chapter 3, I developed the idea for the project in collaboration with Michael Blanton. I implemented the project using catalogs constructed by John Moustakas from observations made by the PRIMUS collaboration (Scott Burles, Alison Coil, Richard Cool, Daniel Eisenstein, Ramin Skibba, Kenneth Wong, and Guangtun Zhu). I wrote the paper with additions from Michael Blanton.

4. For Chapter 4, I developed the idea for the project in collaboration with Jeremy Tinker. I implemented the project using simulation data from Andrew Wetzel. I wrote the paper with comments and edits by Jeremy Tinker and Andrew Wetzel.

Chapter 1

The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles

This Chapter is joint work with Roman Scoccimarro (NYU), Michael R. Blanton (NYU), Jeremy L. Tinker (NYU), and Sergio Rodríguez-Torres (Universidad Autónoma de Madrid) published in the *Monthly Notices of the Royal Astronomical Society* as Hahn et al. (2017).

1.1 Chapter Abstract

Fiber-fed multi-object spectroscopic surveys, with their ability to collect an unprecedented number of redshifts, currently dominate large-scale structure studies. However, physical constraints limit these surveys from successfully collecting redshifts from galaxies too close to each other on the focal plane. This ultimately leads to significant systematic effects on galaxy clustering measurements. Using simulated mock catalogs, we demonstrate that fiber collisions have a significant impact on the power spectrum, $P(k)$, monopole and quadrupole that exceeds sample variance at scales smaller than $k \sim 0.1\ h/\mathrm{Mpc}$.

We present two methods to account for fiber collisions in the power spectrum. The first statistically reconstructs the clustering of fiber collided galaxy pairs by modeling the distribution of the line-of-sight displacements between them. It also properly accounts for fiber collisions in the shot-noise correction term of the $P(k)$ estimator. Using this method, we recover the true $P(k)$ monopole of the mock catalogs with residuals of $< 0.5\%$ at $k = 0.3\ h/\mathrm{Mpc}$ and $< 4\%$ at $k = 0.83\ h/\mathrm{Mpc}$ – a significant improvement over existing correction methods. The quadrupole, however, does not improve significantly.

The second method models the effect of fiber collisions on the power spectrum as a convolution with a configuration space top-hat function that depends on the physical scale of fiber collisions. It directly computes theoretical predictions of the fiber-collided $P(k)$ multipoles and reduces the influence of smaller scales to a set of nuisance parameters. Using this method, we reliably model the effect of fiber collisions on the monopole and quadrupole down to the limiting scales of theoretical predictions. The methods we present in this paper will allow us to robustly analyze galaxy power spectrum multipole measurements to much smaller scales than previously possible.

1.2 Introduction

Cosmological measurements such as galaxy clustering statistics are no longer dominated by uncertainties from statistical precision, but from systematic effects of the measurements. This is a result of the millions of redshifts to distant galaxies that have been obtained through redshift surveys such as the 2dF Galaxy Redshift Survey (2dFGRS; Colless 1999) and the Sloan Digital Sky Survey III Baryon Oscillation Spectroscopic Survey (SDSS-III BOSS; Anderson et al. 2012; Dawson et al. 2013a). Current surveys, such as the Extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2015), and future surveys such as the Dark Energy Spectroscopic Instrument (DESI; Schlegel et al. 2011; Morales et al. 2012; Makarem et al. 2014), and the Subaru Prime Focus Spectrograph (PFS; Takada et al. 2014), will continue to collect many more million redshifts, extending our measurements to unprecedented statistical precision. These completed and future surveys, all use and will use fiber-fed spectrographs.

For each galaxy, a fiber is used to obtain a spectroscopic redshift. However, the physical size of the fiber housing and other physical constraints limit how well any of these surveys can observe close pairs of galaxies. In the SDSS, if two galaxies are located within the fiber collision angular scale from one another on the sky, separate fibers cannot be placed adjacently to observe them simultaneously (Yoon et al. 2008). In these situations, only a single redshift is measured. With redshifts of galaxies in close angular proximity missing from the sample, any clustering statistic probing these scales will be systematically affected.

As our cosmological surveys extend further to higher redshifts, the systematic effect becomes more severe. The fiber collision angular scale corresponds to a larger comoving scale at higher redshift, thereby affecting our measurements on larger scales. BOSS, in particular, has an angular fiber collision scale of 62". This corresponds to ~ 0.43 Mpc/h at the center of the survey's redshift range; fiber-collided galaxies account for $\sim 5\%$ of the galaxy sample (Anderson et al. 2012; Reid et al. 2012; Guo et al. 2012a). While this may seem like a relatively small fraction of redshifts, its effect on clustering measurements such as the power spectrum and bispectrum is significant and needs to be accounted for in order to probe mildly non-linear scales. Unfortunately, future spectroscopic surveys such as DESI, which will use robotic fiber positioner technology, will be subject to similar effects. In fact, based on the DESI Final Design Report[1], which estimates that $\sim 6\%$ of Luminous Red Galaxies and $> 20\%$ of Emission-Line Galaxies will be fiber-collided, fiber collisions

[1] DESI Final Design Report: `http://desi.lbl.gov/tdr/`

will affect a *larger* fraction of the target sample than in BOSS. Therefore, accounting for the effects of fiber collisions will remain a crucial and unavoidable challenge for analyzing clustering measurements.

To correct for fiber collisions, one common approach used in clustering measurements is the nearest neighbor method (Zehavi et al. 2002, 2005, 2011; Berlind et al. 2006a; Anderson et al. 2012). For fiber-collided galaxies without resolved redshifts, the method assigns the statistical weight of the fiber-collided galaxy to its nearest angular neighbor. This provides a reasonable correction for the fiber collision effects at scales much larger than the fiber collision scales; however the correction falls short elsewhere. In fact, as Zehavi et al. (2005) find, fiber collisions affect the two-point correlation function (2PCF) measurements even on scales significantly larger than the fiber collision scale (> 1 Mpc/h).

For power spectrum measurements in BOSS, the nearest neighbor method has recently been supplemented with adjustments in the constant shot-noise term of the power spectrum estimator to correct for fiber collisions (Beutler et al., 2014b; Gil-Marín et al., 2014, 2016a,b; Beutler et al., 2016; Grieb et al., 2016). More specifically, methods like the one used in Gil-Marín et al. (2014) obtain the value of the shot-noise term from mock catalogs and thus rely entirely on their accuracy to correct for fiber collisions. This is concerning since, as we shall demonstrate in detail, fiber collisions depend systematically on the small-scale power spectrum, and mock catalogs used for large scale structure analyses are typically not based on high resolution N-body simulations. In addition, there is no way to validate and calibrate the shot-noise term independently for observations. A more reliable approach is to marginalize over the value of the shot-noise term, and this is the approach that has recently become more popular (Gil-Marín et al., 2016a; Beutler et al., 2016; Grieb et al., 2016; Gil-Marín et al., 2016b). However, adjustments to the shot-noise term are limited to the power spectrum monopole, since higher order multipoles do not have a shot-noise term. However, as we shall discuss in detail below, *fiber collisions affect all multipoles in a k-dependent way*, not just adding a constant for the monopole power.

Guo et al. (2012a), focusing on SDSS-III BOSS like samples, proposed a fiber collision correction method for the 2PCF that is able to reasonably correct for fiber collisions above and below the collision scale. Guo et al. (2012a) estimates the total contribution of fiber-collided galaxies to the 2PCF by examining the pair statistics in overlapping tiling regions of the survey, where a smaller fraction of galaxies suffer from fiber collisions. Unfortunately, applying an analogous method in Fourier space proves to be more difficult. The Guo et al. (2012a) method in Fourier space would involve measuring the power spectra for individual overlapping regions. Given the complex geometry of these regions, the systematic effect introduced by the window function makes measuring the power spectrum at larger scales intractable.

Meanwhile, galaxy redshift-space power spectrum models from perturbation theory continue to reliably model higher k in the weakly non-linear regime (Taruya et al., 2010, 2014; Okumura et al., 2015; Beutler et al., 2016; Grieb et al., 2016; Sanchez et al., 2016). Recent analyses of galaxy power spectrum multipoles (Beutler et al. 2014b; Gil-Marín et al. 2014, 2016a,b; Beutler et al. 2016; Grieb et al. 2016) use scales up to $k_{\max} = 0.15 - 0.2h/\mathrm{Mpc}$ for BOSS galaxies, and this limit will for sure move towards smaller scales in upcoming analyses. As statistical errors decrease the importance of systematics due to fiber collisions plays an increasingly important role. The main goal of this paper is to quantify this systematic effect for the power spectrum multipoles and to provide ways to overcome it; for this purpose we develop two distinct approaches.

The first approach improves upon the nearest neighbor method by modeling the distribution of the line-of-sight displacement between resolved fiber collided galaxies to statistically reconstruct the clustering of fiber-collided galaxies. This uses information on resolved fiber collided galaxies that is available from the data themselves (e.g. in tiling overlap regions). The difficulty with this method is that it works statistically, i.e. we cannot reconstruct the *actual* galaxy by galaxy line of sight displacement due to collisions. As a result of this, while the method works very well to recover the true power spectrum monopole from fiber collided galaxy catalogs, it does not work sufficiently well for the power spectrum quadrupole which is far more sensitive to the precise structure of “fingers of god”.

The second approach addresses the shortcomings of the first one by modeling the effects of fiber collisions on the *predictions* instead of trying to undo their effect on the data before computing power spectrum statistics. It approximates the effect of fiber collisions on the 2PCF as a 2D top hat function. Then it derives the effect of fiber collisions on the galaxy power spectrum as a convolution of the true power spectrum with the top hat function. Therefore the theoretical predictions for the power spectrum are fiber collided and then can be compared directly to the observed fiber collided power spectrum in clustering analyses.

This paper is organized as follows. In Section 1.3, we briefly describe the simulated mock catalogs with realistic fiber collisions and the power spectrum estimator used throughout the paper. We then demonstrate the impact of fiber collisions on power spectrum measurements and how the nearest neighbor method does not adequately account for fiber collisions in Section 1.4.1. We present our two methods of accounting for fiber collisions along with the results for mock catalogs in Section 1.4.2 and Section 1.4.3, respectively. Finally in Section 4.8 we summarize our results and conclude.

1.3 Fiber-collided Mock catalogs

For various purposes, such as characterizing the impact of the survey window function on statistics and estimating covariance matrices, simulated mock catalogs play a crucial role in interpreting clustering measurements of observed galaxies (Cole et al., 1998; Scoccimarro & Sheth, 2002; Anderson et al., 2012; Beutler et al., 2014b; Carretero et al., 2015, also see citations in Chuang et al. 2015b). They also provide a means of understanding systematic effects such as fiber collisions (Guo et al. 2012a; Manera et al. 2013). Since systematic effects can be simulated on them, they allow us to test how these effects influence clustering measurements and devise correction methods that attempt to account for these effects.

A direct way of understanding the effects of fiber collisions on clustering statistics in observations is to first apply fiber collisions to mock catalogs and then compare the clustering statistics obtained from mock catalogs with and without the fiber collisions. Correction methods for fiber collisions can then be applied to the fiber-collided mocks. The merit of the correction method can be assessed by how successfully they reproduce the clustering statistics of the original mock catalogs without fiber collisions. The correction method can then be applied to the observed data with some assurance that it accounts for fiber collisions and improves the clustering measurements.

When applying the fiber collisions to the mock catalogs, it is essential to apply them in the same manner they affect the observations. For BOSS, galaxies within 62” are fiber-collided (Anderson et al. 2012). In reality, this criteria is further complicated by the tiling scheme of observing plates that create overlapping regions, which have a higher success rate in resolving galaxy spectra within the fiber collision angular scale (Guo et al. 2012a; Reid et al. 2012). Furthermore, fiber collisions are only one of the systematic effects that influence BOSS data. Systematic effects include the unique geometry of the BOSS survey, the variable completeness in different areas covered by unique sets of spectroscopic plates, and redshift failures (Anderson et al. 2012; Ross et al. 2012).

Effects of fiber collisions must be understood and interpreted in conjunction with the other systematic effects. Therefore, in this paper, we use Quick Particle Mesh (White et al.

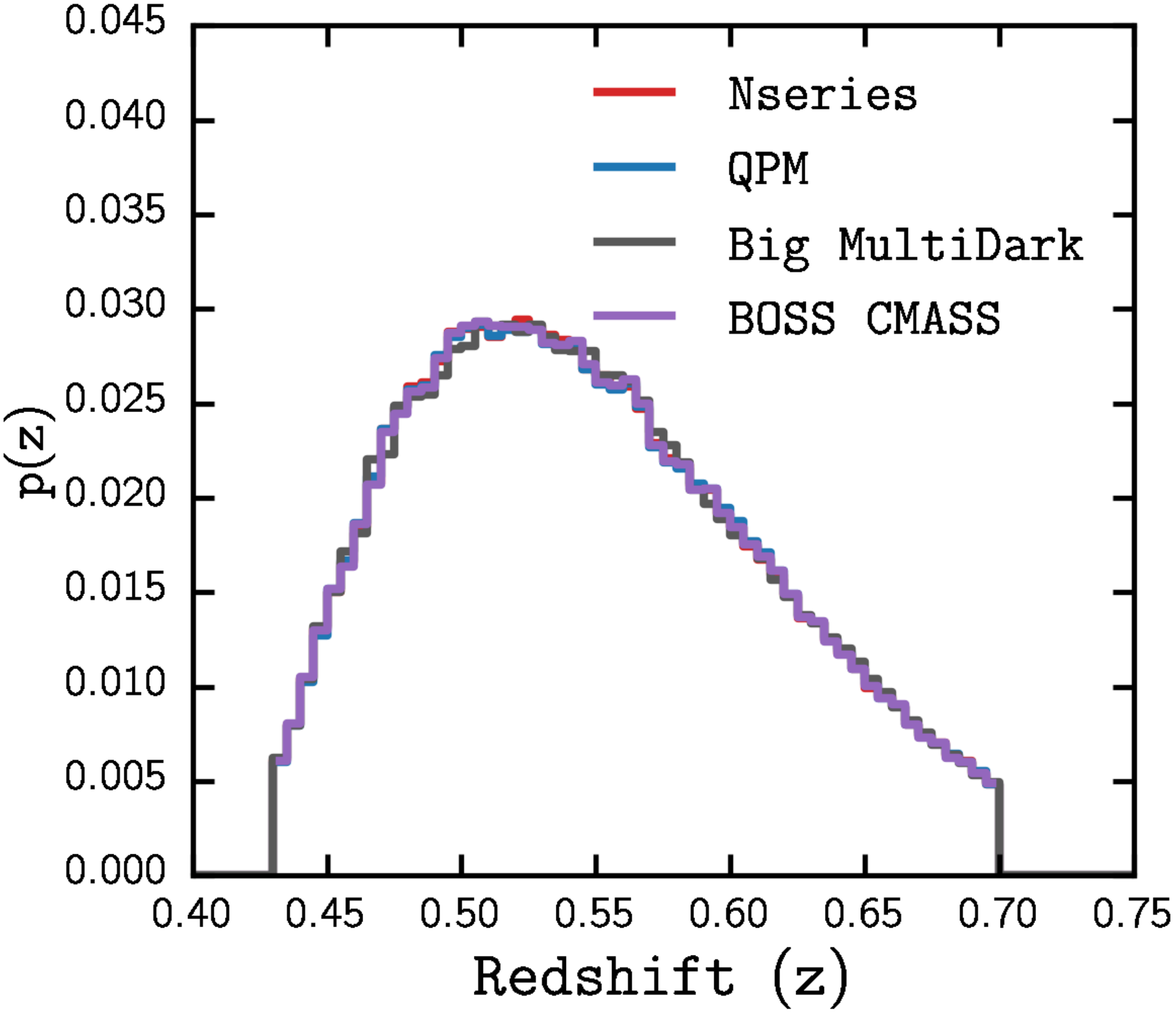

Figure 1.1: Normalized galaxy redshift distribution of the Nseries (orange), QPM (blue), and BigMultiDark (red) mock catalogs. The normalized redshift distribution of BOSS DR12 CMASS sample galaxies is also plotted (black). Each of the distributions were computed with a bin size of $\Delta z = 0.025$. All of the mock catalogs used in this work closely trace the BOSS CMASS redshift distribution.

2014), Nseries (Tinker et al. in prep), and the BigMultiDark (Rodríguez-Torres et al. 2015) mock catalogs, which have already been extensively used in interpreting clustering results for BOSS and are generated through different prescriptions. Therefore they provide a robust sets of data to measure the effects of fiber collisions and to test our correction methods.

The QPM mock galaxy catalogs uses a “quick particle mesh” method, which uses a low resolution particle-mesh N-body solver, with a resolution of 2 Mpc/h, to evolve particles within a periodic simulation volume. The particles are assigned halo masses in order to match the halo mass function and large-scale bias of halos of high resolution simulations. Afterwards the HOD parameterization of Tinker et al. (2012) is used to populate the halos. The mock galaxy sample is then trimmed to the BOSS CMASS survey footprint, downsampled based on angular sky completeness (sector completeness) and radial selection. Furthermore, QPM mocks model the fiber collisions of the BOSS CMASS sample (62”). QPM uses the following ΛCDM cosmology: $\Omega_\mathrm{m} = 0.29$, $\Omega_\Lambda = 0.71$, $\sigma_8 = 0.8$, $n_\mathrm{s} = 0.97$ and $h = 0.7$. We use 100 realizations of the QPM catalog. For a detailed description of the QPM galaxy mock catalogs we refer readers to White et al. (2014).

Next, the Nseries mock catalogs are created from a series of high-resolution N-body simulations. Each mock has the same angular selection function as the North Galactic Cap region of the BOSS DR12 large-scale structure sample for CMASS galaxies (Cuesta et al. 2016b). They also reproduce the redshift distribution of the BOSS CMASS sample. The Nseries mock catalogs are created from seven independent N-body simulations, each of the same cosmology. Each simulation box is 2.5 Gpc/h per side with cosmology: $\Omega_\mathrm{m} = 0.286$, $\Omega_\Lambda = 0.714$, $\sigma_8 = 0.82$, $n_\mathrm{s} = 0.96$ and $h = 0.7$. Out of these Nseries box simulations, the three orthogonal projections of each box is used to create 84 mocks. Each of the cut-out mocks is then passed through the same fiber assignment code as the actual BOSS data using the distribution of plates in BOSS. Thus, the angular variation of fiber collisions faithfully reproduces that of the data, with $\sim 5\%$ of the targets without fibers due to close neighbors in regions of the footprint only covered by one tile.

Finally the BigMultiDark galaxy mock catalog is generated using the BigMultiDark Planck (BigMDPL), one of the MultiDark3 N-body simulations (Klypin et al. 2014). BigMDPL uses a GADGET-2 code (Springel 2005) in a cubic box of 2.5 h^{-1}Gpc sides with 3840^3 dark matter particles and a mass resolution of $2.4 \times 10^{10} h^{-1} M_\odot$. As the name suggests, BigMDPL uses Planck cosmological parameters in a flat ΛCDM cosmology: $\Omega_m = 0.307$, $\Omega_B = 0.048$, $\Omega_\lambda = 0.693$, $\sigma_8 = 0.829$, $n_s = 0.96$ and $h = 0.678$.

From the BigMDPL N-body simulation, Rodríguez-Torres et al. (2015) uses the `RockStar` (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement) halo finder (Behroozi et al. 2013c) to obtain a dark matter halo catalog. Afterwards, they use the SUrvey GenerAtoR code (`SUGAR`) to generate a galaxy catalog from the halo catalog. `SUGAR` uses halo abundance matching with an intrinsic scatter on the stellar mass function of the Portsmouth SED-fit DR12 stellar mass catalog (Maraston et al. 2013) to populate the dark matter halos with galaxies. Rodríguez-Torres et al. (2015) then model fiber collisions using Guo et al. (2012a) in order to reproduce the effect of fiber collisions on the observed BOSS galaxies. For any further details on the BigMultiDark galaxy mock catalog, we refer readers to Rodríguez-Torres et al. (2015).

In Figure 1.1, we plot the normalized redshift distribution of the Nseries (orange), QPM (blue), and BigMultiDark (red) mock catalogs along with the redshift distribution of the BOSS DR12 CMASS sample galaxies. All of these mock catalogs were constructed for the BOSS analysis and their redshift distributions closely trace the observed BOSS distribution.

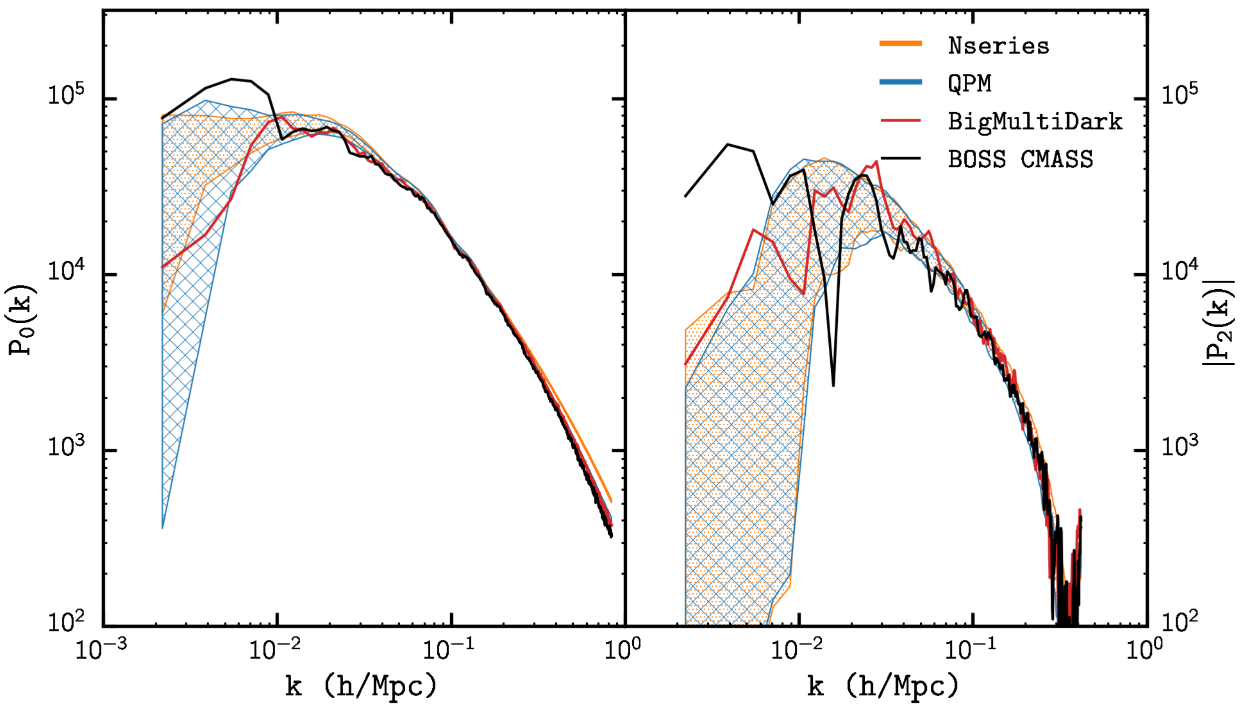

Figure 1.2: Power spectrum monopole $P_0(k)$ and quadrupole $|P_2(k)|$ measurements for the Nseries (orange), QPM (blue), and BigMultiDark (red) mock catalogs (Section 1.3). The $P_l(k)$ measurements for the Nseries and QPM mock catalogs are averaged over the multiple mock realizations and the width of the power spectra represents the sample variance ($\sigma_l(k)$; Eq. 1.10) of the realizations. For the quadrupole, we plot the $|P_2(k)|$ instead of $P_2(k)$ because the measurement becomes negative for $k \gtrsim 0.35$ h/Mpc. For comparison, we also include the monopole and quadrupole power spectra of the BOSS DR12 CMASS sample, which are calculated using the same estimator but with statistical weights described in Eq. (1.9). While fiber collisions are inevitably included in the BOSS CMASS power spectra, they are *not* yet applied to the mock catalogs power spectra measurements above.

1.3.1 Power Spectrum Estimator

In this paper, out of the many possible clustering measurements, we focus on the galaxy power spectrum and its monopole and quadrupole in redshift space. Throughout the paper, unless specified, when we measure the power spectrum we use the estimator described in Scoccimarro (2015), which accounts for radial redshift space distortions (see also Bianchi et al. 2015). In this estimator, galaxies are interpolated and Fast Fourier transformed as discussed in Sefusatti et al. (2016). Since the algorithm is efficient, it makes power spectrum computations for large number of mock realizations tractable.

To summarize the method, we calculate the monopole component of the power spectrum using:

$$\widehat{P_0}(k) = \frac{1}{I_{22}} \left[\int \frac{d\Omega_k}{4\pi} |F_0(\mathbf{k})|^2 - N_0 \right] \tag{1.1}$$

where

$$F_0(\mathbf{k}) = \left(\sum_{j=1}^{N_g} - \alpha \sum_{j=1}^{N_r} \right) w_j \; e^{i\mathbf{k}\cdot\mathbf{x}_j} \tag{1.2}$$

with normalization constant

$$I_{22} = \alpha \sum_{j=1}^{N_r} \bar{n}(\mathbf{x}_j) w_j^2 \tag{1.3}$$

and shot noise term following from the estimator is (Scoccimarro, 2015)

$$N_0 = \left(\sum_{j=1}^{N_g} + \alpha^2 \sum_{j=1}^{N_r} \right) w_j^2, \tag{1.4}$$

which represents the constant shot noise contribution to the power due to the discrete density field of our galaxies and random catalog. Here α is the ratio of the number of galaxies (N_g) over the number of synthetic random galaxies (N_r), $\bar{n}(\mathbf{x})$ is the mean density of the galaxies at position $\mathbf{x}$, and w_j is weight of each object, which includes the minimum variance weight from Feldman et al. (1994):

$$w_{\rm FKP}(\mathbf{x}_j) = \frac{1}{1+\bar{n}(\mathbf{x}_j)P_0} \tag{1.5}$$

where P_0 is the power spectrum amplitude at which the error is minimized. We use $P_0 = 20000\ \mathrm{Mpc}^3/h^3$ for our analysis, which corresponds to $k \sim 0.1\ h/\mathrm{Mpc}$. We note that the shot noise term in Eq. (1.4) differs from the standard shot noise term from Feldman et al. (1994). The difference between various shot noise expressions used in the literature will be discussed in detail in Section 1.4.2.2.

For the quadrupole, we have

$$\widehat{P_2}(k) = \frac{5}{I_{22}} \int \frac{d\Omega_k}{4\pi} F_2(\mathbf{k}) F_0^*(\mathbf{k}) \tag{1.6}$$

where

$$F_2(\mathbf{k}) = \frac{3}{2}\hat{k}_a \hat{k}_b Q^{ab}(\mathbf{k}) - \frac{1}{2}F_0(\mathbf{k}) \tag{1.7}$$

with

$$Q^{ab}(\mathbf{k}) = \left(\sum_{j=1}^{N_g} - \alpha \sum_{j=1}^{N_r} \right) \hat{x}_j^a \hat{x}_j^b w_j\ e^{i\mathbf{k}\cdot\mathbf{x}_j} \tag{1.8}$$

In Figure 1.2, we plot the power spectrum monopole and quadrupole, $P_0(k)$ and $|P_2(k)|$, measured using Eq. (1.1) and Eq. (1.6), respectively, for the Nseries, QPM, and BigMultiDark mock catalogs. We plot $|P_2(K)|$ because the power spectrum quadruple becomes negative for $k \gtrsim 0.35\ h/\mathrm{Mpc}$. $P_0(k)$ and $|P_2(k)|$ are averaged over the 84 and 100 realizations for Nseries and QPM. We note that fiber collisions are not applied to these mock catalogs. Without fiber collisions, the weights of the objects are equivalent to the FKP weights, $w_j = w_{j,\rm FKP}$.

We also plot the $P_0(k)$ and $P_2(k)$ of the BOSS Data Release 12 CMASS data (black) in Figure 1.2. For BOSS DR12 CMASS, systematic weights are assigned to the galaxies in order to account for sector completeness, redshift failures, and fiber collisions. Each galaxy has a statistical weight determined by,

$$w_{j,\mathrm{tot}} = w_{j,\mathrm{sys}}(w_{j,\mathrm{rf}} + w_{j,\mathrm{fc}} - 1), \tag{1.9}$$

(Anderson et al. 2012; Ross et al. 2012; Beutler et al. 2014b), which are included in the final object weight w_j along with $w_{j,\mathrm{FKP}}$. In this formula, $w_{j,\mathrm{rf}}$ is a weight that accounts for redshift failures and $w_{j,\mathrm{fc}}$ is the fiber collision weight determined by the nearest angular neighbor method, which we later discuss in Section 1.4.1. The statistical weights are also included in $\alpha = \sum_{j=1}^{N_g} w_{\mathrm{tot}}/N_r$. We note that fiber collisions are inevitably included in the CMASS $P_l(k)$. However they are not yet included in the $P_l(k)$ of the mock catalogs in Figure 1.2.

For the mock catalogs with multiple realizations (QPM and Nseries), we compute the sample variance of the power spectrum

$$\sigma_l(k) = \sqrt{\frac{1}{N_{\mathrm{mocks}} - 1} \sum_{i=1}^{N_{\mathrm{mocks}}} \left(P_l^i(k) - \langle P_l(k) \rangle\right)^2}\ . \tag{1.10}$$

N_{mock} is the number of mock realizations (84 for Nseries and 100 for QPM) and $P_l^i(k)$ is the power spectrum for each realization. $\sigma_l(k)$ is represented in Figure 1.2 by the width of the shaded regions.

1.4 Fiber Collision Methods

1.4.1 Nearest Angular Neighbor Method (NN)

A common approach to accounting for fiber collisions in clustering measurements has been to use the nearest angular neighbor method (Zehavi et al. 2002, 2005; Berlind et al.

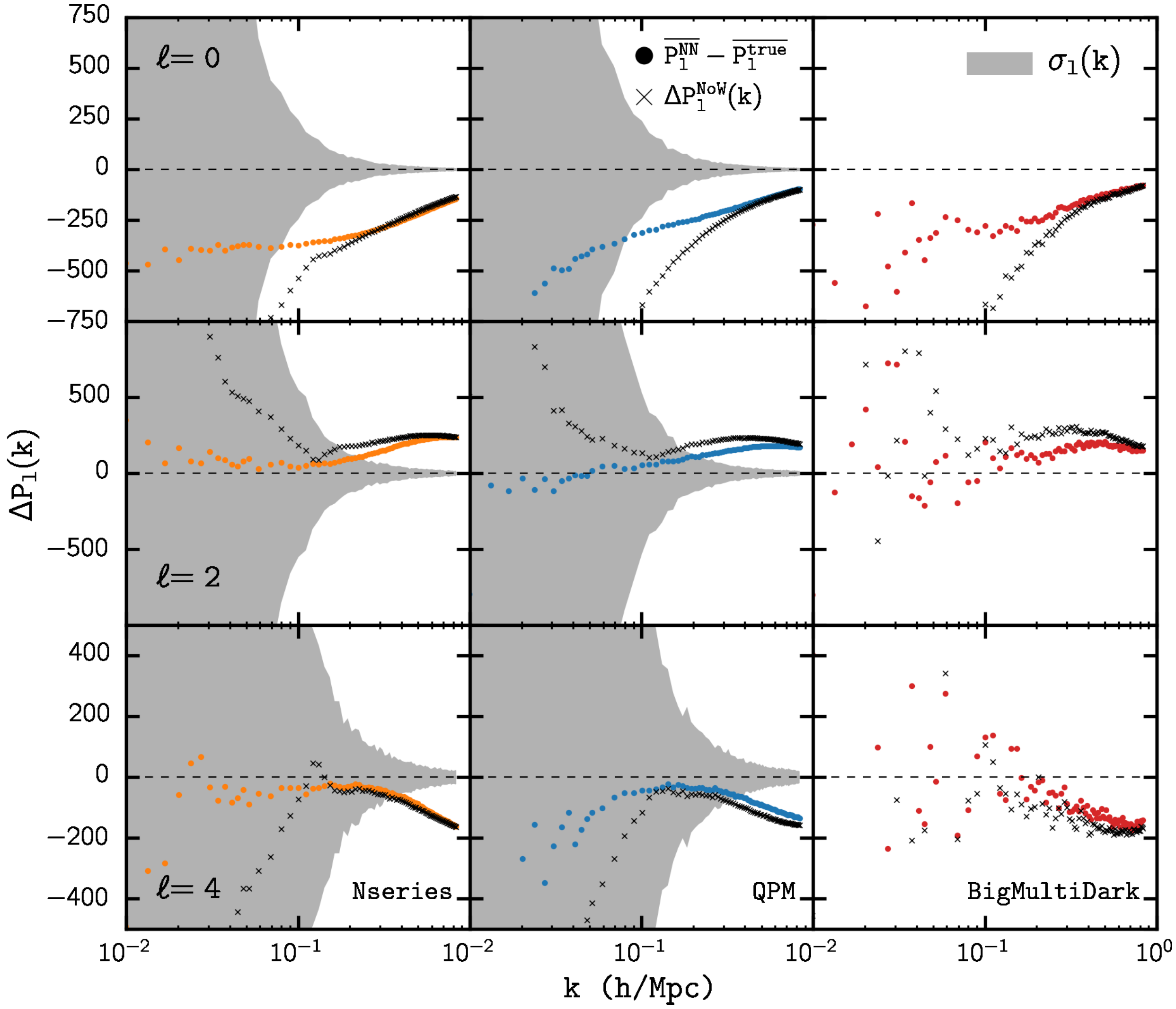

Figure 1.3: The fiber collision power spectrum residual, $(P_l^{\rm NN} - P_l^{\rm true})$ (Section 1.4.1), for the monopole (top), quadrupole (middle), and hexadecapole (bottom) of the Nseries (left), QPM (middle), and BigMultiDark (right) mock catalogs. For the Nseries and QPM mocks, we plot the sample variances $\sigma_l(k)$ (grey shaded region) of $P_l^{\rm true}(k)$ for comparison. The power spectrum residual for the NN method is an improvement over the residual with no correction ($\Delta P_l^{\rm NoW}(k)$; x) at most scales probed. However, we highlight that at $k > 0.1$ h/Mpc and $k > 0.2$ h/Mpc, for the monopole and quadrupole respectively, the residuals from fiber collision surpass the sample variance. At smaller scales, NN method does not sufficiently account for the effects of fiber collisions in $P_l(k)$ measurements.

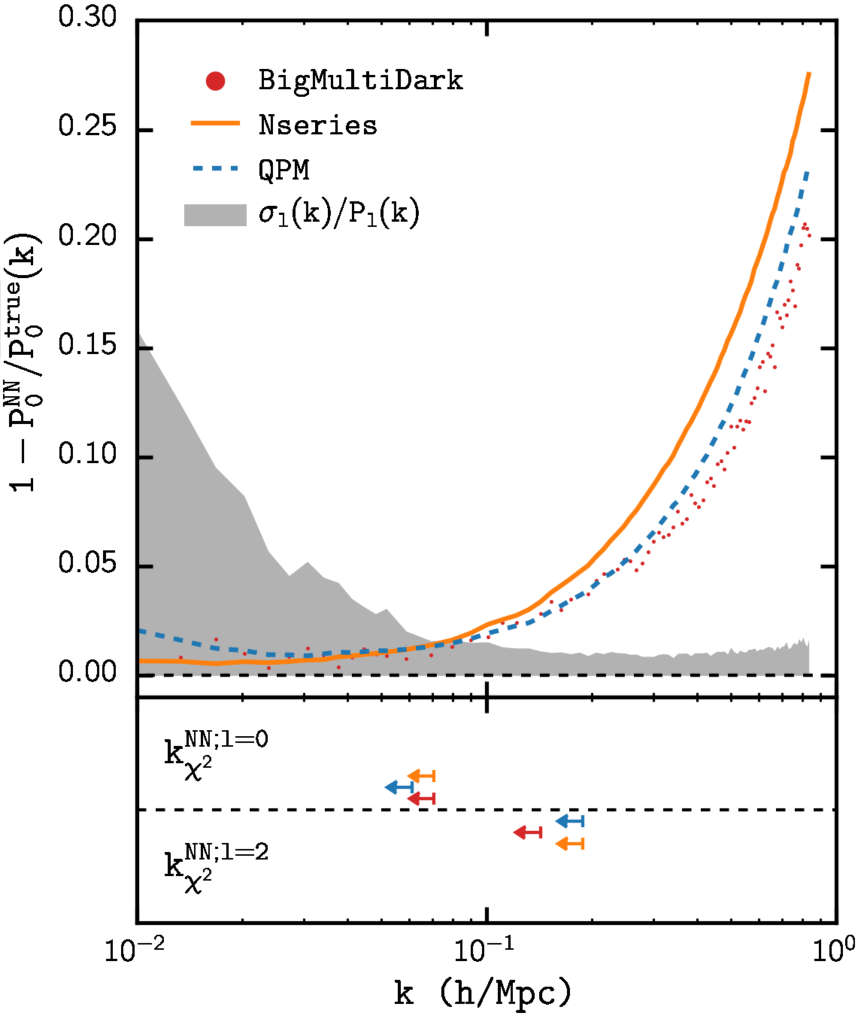

Figure 1.4: *Top Panel*: The normalized residuals, $1-\overline{P_0^{\rm NN}}/\overline{P_0^{\rm true}}(k)$, of the NN method for the Nseries (orange), QPM (blue), and BigMultiDark (red) power spectrum monopole. We also plot the normalized sample variance $\sigma_0(k)/P_0(k)$ (gray shaded region) of the Nseries mocks for comparison. The QPM $\sigma_0(k)/P_0(k)$ is effectively the same as the Nseries $\sigma_0(k)/P_0(k)$, so we do not included in the figure. The comparison reveals that the effect of fiber collisions not only biases the power spectrum beyond sample variance at $k \gtrsim 0.1$ h/Mpc, but that the effect increases relative to sample variance at smaller scales. At $k = 0.2$ h/Mpc, the normalized residual is greater than 4 times the normalized sample variance. *Bottom Panel*: We mark k_{χ^2} where $\Delta\chi^2(k_{\chi^2}) = 1$ (Eq. 1.11) for the NN method. $k_{\chi^2}^{\rm NN}$ is a conservative scale limit of the NN method. Arrows above the dashed line mark k_{χ^2} for the monopole while the arrows below the dashed line mark k_{χ^2} for the quadrupole. The color of the arrows indicate the mock catalog: Nseries (orange), QPM (blue), and BigMultiDark (red). Averaged over the three mock catalogs, we get $k_{\chi^2}^{\rm NN} = 0.068$ and 0.17 h/Mpc. for the monopole and quadrupole respectively.

2006a; Zehavi et al. 2011; Anderson et al. 2012), hereafter NN method. For galaxies without resolved spectroscopic redshifts due to fiber collisions, the entire statistical weight of the galaxy is assigned to its nearest angular neighbor with resolved redshift. According to Zehavi et al. (2002), this method effectively assumes that all galaxies within the angular fiber collision scale (< 62" for BOSS) are correlated with one another. In the context of the halo model, the NN method assumes that galaxies within the fiber collision angular scale reside in the same halo so displacing one of the galaxies and placing it on top of the other does not significantly impact clustering statistics. This is a reasonable assumption for the 2PCF and the power spectrum on scales far greater than fiber collisions.

One consequence of this method is that galaxies coincidentally within the angular fiber collision scale (hereafter referred to as "chance alignments") are incorrectly assumed to be gravitationally correlated and within the same halo. So when the statistical weight of the collided galaxy is added to its nearest angular neighbor, the collided galaxy is in fact displaced significantly from its true radial position. This displacement can even be on the scale of the survey depth, which corresponds to ~ 500 Mpc for BOSS. Furthermore, even for fiber collided galaxies that reside in the same gravitationally bound structures such as groups or clusters, up-weighting the nearest neighbor disregards the line-of-sight displacements within these structures.

To precisely quantify the effect of fiber collisions on the power spectrum, we compare the power spectrum measurements of the NN weighted fiber collided mock catalogs P_l^{NN} to the power spectrum measurements of the mock catalogs without fiber collisions, the "true" power spectrum P_l^{true}. Specifically, in Figure 1.3, we plot the power spectrum residual ($P_l^{\mathrm{NN}} - P_l^{\mathrm{true}}$) as a function of k. The power spectrum estimators Eq. (1.1) and (1.6) are used to calculate the monopole (top) and quadrupole (center) respectively. We include measurements of the sample variance, $\sigma_l(k)$, for the Nseries and QPM mock catalogs (Eq. 1.10). We also include the power spectrum residual $\Delta P_l^{\rm NoW}(k) = P_l^{\rm NoW}(k) - P_l^{\rm true}(k)$ (dashed), where $P_l^{\rm NoW}(k)$ is the power spectrum of the fiber collided mock catalogs with *no* NN weights, with the collided galaxies removed from the sample. In this paper we focus on the monopole and quadrupole, however for reference, we also include the effect of fiber collisions on the power spectrum hexadecapole (bottom).

As both $P_l(k)$ and $\sigma_l(k)$ vary significantly over the probed k range, the significance of the discrepancies between $P_l^{\rm NN}(k)$ and $P_l^{\rm true}(k)$ are not adequately portrayed in Figure 1.3, especially for the monopole. Therefore, to compare $P_0^{\rm NN}$ and $P_0^{\rm true}$ over a wide k range and to especially highlight the discrepancies at small scales, in Figure 1.4, we compare the normalized monopole residuals, $1 - P_0^{\rm NN}/P_0^{\rm true}$, to the normalized sample variance, $\sigma_0(k)/P_0^{\rm true}$.

For the monopole, Figure 1.3 demonstrates that while the NN method (circles) provides an overall improvement over applying no correction (crosses) at most scales, fiber collisions still significantly bias the corrected power spectrum at all scales. The effect also has a significant k dependence, which implies that an adjusted constant shot noise term alone is insufficient in accounting for the deviation. Even at $k \approx 0.1$ h/Mpc, the effect of fiber collisions in the NN method alarmingly surpasses sample variance. While the amplitude of the residual decreases as k increases, Figure 1.4 reveals that as a fraction of $P_0^{\rm true}(k)$, the discrepancy is in fact increasing. In other words, the NN method becomes less effective at correcting for fiber collisions on smaller scales, as expected. At the smallest scales probed ($k = 0.83$ h/Mpc), the $P_0^{\rm NN}(k)$ underestimates the true power spectrum monopole by over 20%.

For the quadrupole, the NN method improves the power spectrum residuals over no correction. However, even with the NN method, the effect of fiber collisions begins to significantly grow at $k = 0.1$ h/Mpc and becomes comparable to the sample variance at $k \sim 0.2$ h/Mpc. For $k > 0.2$ h/Mpc, the effect continues to increase and quickly overtakes the decreasing sample variance. At the smallest scales measured ($k = 0.83$ h/Mpc) the residual is over eight times the sample variance.

Recently power spectrum analyses have measured the power spectrum using a wide range of k bins: for example, Anderson et al. (2012) use $\Delta k = 0.04$ h/Mpc and Beutler et al. (2014b) and Grieb et al. (2016) use $\Delta k = 0.005$ h/Mpc. Here, we use $\Delta k = 0.01$ h/Mpc, which is within this general range, in agreement with Beutler et al. (2016) and Gil-Marín et al. (2016a). Sample variance measured with larger Δk is smaller; so a straight comparison in Figure 1.3 between the power spectrum residuals (symbols) and the sample variance (shaded region) has a significant dependence on the choice of Δk. What is independent of binning is a cumulative χ^2 as a function of k, and thus we define a k scale limit k_{χ^2} so that $\Delta\chi^2(k_{\chi^2}) = 1$, where

$$\Delta\chi^2(k') = \sum_{i,j<N_k} \left[P^{\mathrm{NN}}_{l,i} - P^{\mathrm{true}}_{l,i}\right] C^{-1}_{l;\,i,j} \left[P^{\mathrm{NN}}_{l,j} - P^{\mathrm{true}}_{l,j}\right] \tag{1.11}$$

where N_k is the number of bins where $k < k'$ and $C^{-1}_{l;i,j}$ are the elements of the inverse covariance matrix for $P^{\mathrm{true}}_l(k)$. The elements of the covariance matrix $\mathbf{C}_l$ are computed as

$$\mathrm{C}_{l;\,i,j} = \frac{1}{N_{\mathrm{mocks}} - 1} \sum_{k=1}^{N_{\mathrm{mocks}}} \left[P^{(k)}_{l;\,i} - \overline{P}_{l;\,i}\right]\left[P^{(k)}_{l;\,j} - \overline{P}_{l;\,j}\right]$$

for the Nseries and QPM mocks. For BigMD, which only has one realization, we use the covariance matrix of the Nseries realizations. In the lower panel of Figure 1.4, we mark the monopole and quadrupole $k^{\mathrm{NN}}_{\chi^2}$ for the mock catalogs using the NN method. Arrows above the dashed line mark the monopole $k^{\mathrm{NN}}_{\chi^2}$ for Nseries (orange), QPM (blue) and BigMultiDark (red) catalogs. Similarly, the arrows below the dashed line mark the quadrupole $k^{\mathrm{NN}}_{\chi^2}$ for the mock catalogs. Averaged over the three mock catalogs, we get $k^{\mathrm{NN}}_{\chi^2} = 0.068$ and 0.17 h/Mpc for the monopole and quadrupole respectively.

At $k = 0.2$ h/Mpc, the fiber collision residual for the monopole is over four times sample variance with average normalized residual of 4.4% compared to the 0.9% normalized sample variance. Moreover, we find that $k_{\chi^2} = 0.068$ h/Mpc, which is well below the maximum wavenumbers used typically in analyses. For the quadrupole, the fiber collision residual is approximately equivalent to sample variance at $k = 0.2$ h/Mpc and $k_{\chi^2} = 0.17$ h/Mpc, but it quickly deteriorates with increasing k. Therefore for theoretical predictions that attempt to go beyond these scales, the effects of fiber collisions undoubtedly dominate the sample variance for both the power spectrum monopole and quadrupole and the NN method proves to be insufficient. In order to correct for this effect, we next present our first approach: the 'line-of-sight reconstruction' method.

1.4.2 Line-of-Sight Reconstruction Method

1.4.2.1 Line-of-Sight Displacement of Fiber Collided Pairs

It is impossible to determine definitively from observed galaxy data whether individual fiber collided galaxies without resolved spectroscopic redshifts are correlated or chance alignments. However, the line-of-sight displacement of fiber collided galaxy pairs with resolved redshifts make it possible to model the overall impact fiber collisions have on displacing galaxies.

For the BOSS galaxy catalog, fiber collided pairs with resolved spectroscopic redshifts are mainly located in the overlapping regions (Section 1.3). For the simulated mock catalogs, fiber collisions are post-processed after the galaxy positions are generated. Therefore, all galaxies in fiber collided pairs have resolved redshifts. From these resolved redshifts we calculate the comoving line-of-sight displacement (d_{LOS}) by taking the difference between

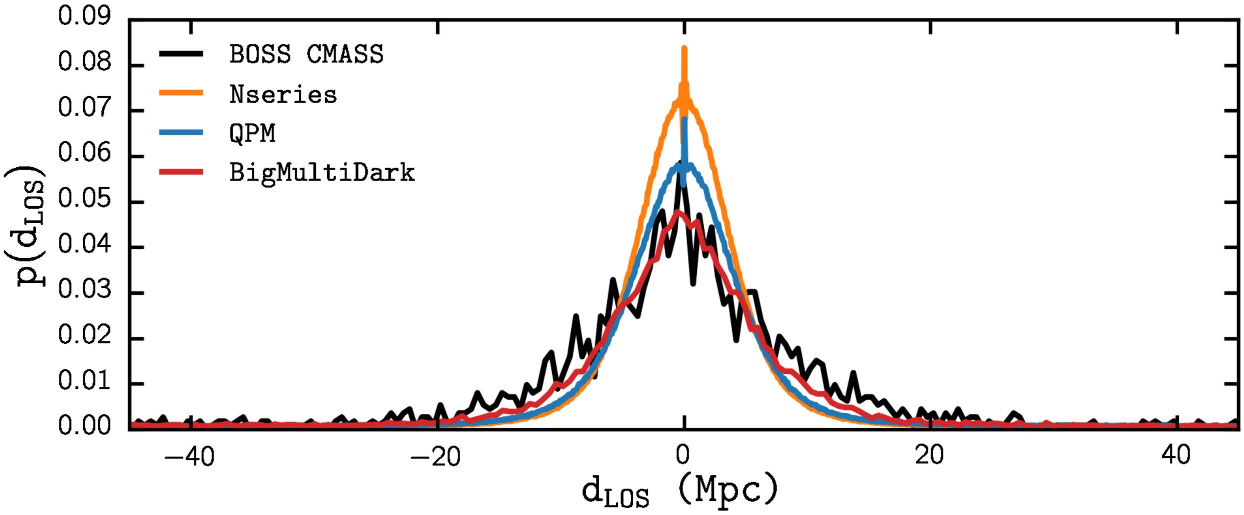

Figure 1.5: Normalized distribution of $d_{\rm LOS}$ for Nseries (orange), QPM (blue), and BigMultiDark (green) mock catalogs. The normalized $d_{\rm LOS}$ distribution of BOSS DR12 is also plotted (black). The mock catalog distributions have bin sizes of $\Delta d = 0.2\,\rm Mpc$, while the CMASS distribution has a bin size of $\Delta d = 0.5\,\rm Mpc$. The distribution extends beyond the range of the above plot to $\sim \pm 500$ Mpc. In the discussion of Section 1.4.2, we focus mainly on the peak of the distribution at roughly -20 Mpc $< d_{\rm LOS} <$ 20 Mpc.

the line-of-sight comoving distance of the resolved redshifts:

$$d_{\rm LOS} = D_{\rm C}(z_1) - D_{\rm C}(z_2). \tag{1.12}$$

$D_{\rm C}(z)$ here is the line-of-sight comoving distance at z (Hogg 1999), and z_1 and z_2 represent the resolved redshifts of the two galaxies in the fiber collided pair.

The normalized distributions of the calculated $d_{\rm LOS}$ for all resolved fiber collided pairs are presented in Figure 1.5 for Nseries (orange), QPM (blue), BigMultiDark (red), and BOSS DR12 (black). The $d_{\rm LOS}$ distributions for all catalogs consist of two components: a peak roughly within the range -20 Mpc $< d_{\rm LOS} <$ 20 Mpc and a flat component (hereafter "tail" component) outside the peak that extends to $d_{\rm LOS} \sim \pm 500$ Mpc. The entire range of the distribution is not displayed in Figure 1.5. For BOSS, as mentioned above, the $d_{\rm LOS}$ distribution only reflects the $d_{\rm LOS}$ values from galaxy pairs within the fiber collision angular scale with resolved spectroscopic redshifts, mostly from overlapping regions of the survey.

Galaxies within the same halo, due to their gravitational interactions at halo-scales, are more likely to be in close angular proximity with each other. These galaxies in over-dense regions cause the peak in the $d_{\rm LOS}$ distribution. The "tail" component consists of chance aligned galaxy pairs that happen to be in close angular proximity in the sky.

Focusing on the peak of the distribution, we note that it closely traces a Gaussian functional form. Therefore, we fit

$$p(d_{\rm LOS}) = A\, e^{-d_{\rm LOS}^2/2\sigma_{\rm LOS}^2} \tag{1.13}$$

for an analytic prescription of the $d_{\rm LOS}$ distribution peak as a function of $d_{\rm LOS}$ for each of the mock catalogs. We list the best-fit $\sigma_{\rm LOS}$ obtained by fitting Eq. (1.13) to the $d_{\rm LOS}$ distribution peak using MPFIT (Markwardt 2009) in Table 1.1. The parameter values in

Table 1.1: $d_{\rm LOS}$ Distribution Best-fit Parameters

Catalog	$\sigma_{\rm LOS}$ (Mpc)	$f_{\rm peak}$
Nseries	3.88	0.69
QPM	4.35	0.62
BigMultiDark	5.47	0.60
CMASS	6.56	0.70

Notes: Best-fit parameter $\sigma_{\rm LOS}$ (Eq. 1.13) and peak fraction $f_{\rm peak}$ (Eq. 1.14) for the $d_{\rm LOS}$ distributions in Figure 1.5.

Table 1.1 and Figure 1.5 illustrate that the $d_{\rm LOS}$ distributions for the mock catalogs closely trace the BOSS DR12 distribution, which encourages our use of these mock catalogs in our investigation.

Using the best-fit to the peak of the $d_{\rm LOS}$ distribution, we estimate the fraction of collided pairs that are within the peak as the ratio of pairs with $|d_{\rm LOS}| < 3\sigma_{\rm LOS}$ over the total number of pairs:

$$f_{\rm peak} = \frac{\sum\limits_{|d_{\rm LOS}|<3\sigma_{\rm LOS}} p(d_{\rm LOS})}{N_{\rm pairs}}, \tag{1.14}$$

where $N_{\rm pairs}$ is the total number of fiber collided pairs. $f_{\rm peak}$ roughly corresponds to the fraction of galaxy pairs that are correlated. The $f_{\rm peak}$ values calculated for the mock catalogs are listed in Table 1.1. They are consistent with the BOSS DR12 $f_{\rm peak}$.

For the NN method of the previous section to be entirely correct, the $d_{\rm LOS}$ distribution in Figure 1.5 would have to be a delta function, which is clearly not the case. By simply incorporating the peak of the $d_{\rm LOS}$ distribution, we can significantly improve clustering statistics on small scales. Rather than placing the fiber collided galaxy on top of its nearest angular neighbor as the NN correction does, placing the fiber collided galaxy at a line-of-sight displacement, sampled from the peak of the $d_{\rm LOS}$ distribution, away from its nearest neighbor better reconstructs the galaxy clustering on small scales.

Only $f_{\rm peak}$ of the collided pairs should be displaced, since only $f_{\rm peak}$ of the fiber collided pairs are correlated. Meanwhile, the other $(1 - f_{\rm peak})$ pairs should retain their NN weights since they are uncorrelated. Displacing these galaxies as well according to the tail piece of the $d_{\rm LOS}$ distribution is not desirable because in an object by object basis we do not know which galaxies should actually be in the tail of the distribution, thus we will be making large mistakes in $d_{\rm LOS}$ galaxy by galaxy. In addition, it is difficult to incorporate that these galaxies should be correlated with others and ignoring this modifies large-scale power. In our approach, the remaining $(1 - f_{\rm peak})$ fiber collided pairs are thus kept with their NN weights, and this is reflected in the shot noise correction of our estimator (Eq. 1.4), which in turn makes connection to previous methods in the literature as we now discuss.

1.4.2.2 Shot Noise Corrections

Measurements of the power spectrum are made on observations of discrete distributions of galaxies rather than continuous density fields. The discreteness contributes to the power spectrum. In order to correct for this contribution, galaxies are assumed to be Poisson samplings of the underlying distribution and a shot noise correction term is included in the power spectrum estimator (Peebles, 1980b; Feldman et al., 1994).

The expectation value of the shot noise term takes the following form (Feldman et al., 1994),

$$P_{\rm shot} = \frac{(1+\alpha)\int d^3r\ \bar{n}(\mathbf{r})w^2(\mathbf{r})}{\int d^3r\ \bar{n}^2(\mathbf{r})w^2(\mathbf{r})}. \tag{1.15}$$

Note that for the case of uniform weights (w = const.), constant number density and no random catalog this reduces to the standard shot-noise Poisson correction $P_{\rm shot} = \bar{n}^{-1}$. In practice the integrals in Eq. (1.15) can be written as discrete sums over the synthetic random catalog (Feldman et al., 1994). $\int d^3r\ \bar{n}(\mathbf{r})...$ is computed as $\alpha \sum_{\rm ran} ....$ Then the shot noise term becomes,

$$P^{\rm FKP}_{\rm shot} = \frac{(1+\alpha)\alpha \sum\limits_{\rm random} w^2_{\rm FKP}(\mathbf{r})}{\alpha \sum\limits_{\rm random} \bar{n}(\mathbf{r})\ w^2_{\rm FKP}(\mathbf{r})}. \tag{1.16}$$

This however, represents the expectation value of the shot noise, not the actual value (Hamilton 1997) since all quantities involved are mean values (calculated through the random catalog). To use the full information provided by the data, the shot noise of the galaxies should be computed from the actual galaxy weights, not the randoms. This simply corresponds to taking the self-pairs in the power spectrum estimator, Eq. (1.1), which leads to Eq. (1.4) and we can rewrite here as,

$$P^{\rm Hahn+}_{\rm shot} \quad = \quad \frac{\sum\limits_{\rm galaxy} w^2_{\rm FKP}(\mathbf{r})\, w^2_{\rm tot}(\mathbf{r}) + \alpha^2 \sum\limits_{\rm random} w^2_{\rm FKP}(\mathbf{r})}{\alpha \sum\limits_{\rm random} \bar{n}(\mathbf{r})\, w^2_{\rm FKP}(\mathbf{r})} \tag{1.17}$$

where $\alpha = (\sum_{\rm gal} w_{\rm tot})/N_r$. We emphasize that this is *the* shot noise of the estimator. In other words, if one takes the limit $k \to \infty$, the estimator in Eq. (1.1) will approach this value if no shot-noise subtraction is applied. The systematic effects from completeness, redshift failures and fiber collisions are accounted for through $w_{\rm tot}$ of the observed galaxies. In our case, $w_{\rm tot} = w_{\rm sys}$ for the resolved $f_{\rm peak}$ fraction of galaxies that have been displaced away from their NN positions, while $w_{\rm tot} > w_{\rm sys}$ for the $(1 - f_{\rm peak})$ fraction of galaxies that are deemed to be in the tail of the LOS distribution and are described by NN weights of the galaxies they collided with.

Recent work in the literature of power spectrum analysis modeled the effect of fiber collisions by solely modifying the shot noise term for the NN method (Beutler et al., 2014b;

Gil-Marín et al., 2014). This assumes that the effect of fiber collisions beyond NN weights is to alter the large-scale effective shot noise, and therefore that only the power spectrum monopole is affected since the quadrupole is free of shot noise. Beutler et al. (2014b) supplements the NN method with a shot noise correction term given by,

$$P_{\rm shot}^{\rm B2014} = \frac{\sum\limits_{\rm galaxy} w_{\rm FKP}^2 w_{\rm tot}(\mathbf{r}) w_{\rm sys}(\mathbf{r}) + \alpha^2 \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}{\alpha \sum\limits_{\rm random} \bar{n}\, w_{\rm FKP}^2(\mathbf{r})}. \tag{1.18}$$

Note that in the first term of the numerator in this equation $w_{\rm fc}$ is only included in $w_{\rm tot}$ as it does not enter in $w_{\rm sys}$. We note that beyond their choice of Eq. (1.18) for the shot noise correction term, Beutler et al. (2014b) marginalizes over a constant stochasticity term in their analysis (see Eq. 40 in Beutler et al., 2014b). Thus, the impact of this particular choice is not straightforward.

Meanwhile, Gil-Marín et al. (2014) constructs $P_{\rm shot}$ using two separate components: one for "true pairs" and the other for "false pairs". The shot-noise contribution to the power from "true pairs" is the same as Eq. (1.18) while the "false pairs" shot-noise contribution is (same as Eq. 1.17),

$$P_{\rm shot}^{\rm False} = \frac{\sum\limits_{\rm galaxy} w_{\rm FKP}^2 w_{\rm tot}^2(\mathbf{r}) + \alpha^2 \sum\limits_{\rm random} w_{\rm FKP}^2(\mathbf{r})}{\alpha \sum\limits_{\rm random} \bar{n}\, w_{\rm FKP}^2(\mathbf{r})}. \tag{1.19}$$

Gil-Marín et al. (2014) calculates the total $P_{\rm shot}$ as the weighted combination of $P_{\rm shot}^{\rm True}$ and $P_{\rm shot}^{\rm False}$:

$$P_{\rm shot}^{\rm GM2014} = (1 - x_{\rm PS}) P_{\rm shot}^{\rm True} + x_{\rm PS}\, P_{\rm shot}^{\rm False} \tag{1.20}$$

In their analysis, Gil-Marín et al. (2014) use $x_{\rm PS} = 0.58$, which they infer by measuring the difference between the true and the fiber-collided power spectrum monopole in the `PTHalos` galaxy mock catalogs (Manera et al. 2013). Unfortunately, since the true power spectrum is the measurement we are trying to recover from the observations, the x_{PS} parameter cannot be inferred or validated from the actual BOSS observations. Moreover, one might worry about relying `PTHalos` or similar methods that are not based on high resolution N-body simulations, to extract corrections for fiber collisions that depend on small-scale power. An extension of this approach is used in recent BOSS analyses (Beutler et al., 2016; Grieb et al., 2016; Gil-Marín et al., 2016b) where Eq. (1.20) is used and is supplemented with a marginalization over the shot noise value. However, as we discussed above, this has no effect in the quadrupole power spectrum, which remains the same as in the NN method.

At this point it is worth casting our "line-of-sight reconstruction" (LRec) method in similar language to the methods we just discussed. We treat the "true pairs" (what we called peak-pairs) by displacing them according to the peak LOS distribution, which modifies all the power spectrum multipoles, and use the NN method for the "false pairs" (pairs in the tail of the LOS distribution). Our shot noise correction is not adjusted, rather it is the true shot noise from the estimator. We now discuss the implementation and performance of our LRec fiber collision method.

1.4.2.3 In Practice

We first begin with fiber collided mock catalogs with the NN fiber collision weights that accurately simulate the effects of fiber collisions on the actual BOSS observations. From this catalog, we construct the d_{LOS} distribution, as described in Section 1.4.2.1 and fit for the best-fit parameters σ_{LOS} and f_{peak} of Eq. (1.13).

We select f_{peak} of the fiber collided galaxy pairs in the catalog and designate them as correlated pairs that lie within the peak of the d_{LOS} distribution. We refer to these fiber collided pairs as "peak-assigned". At this point, each of these pairs, based on their NN

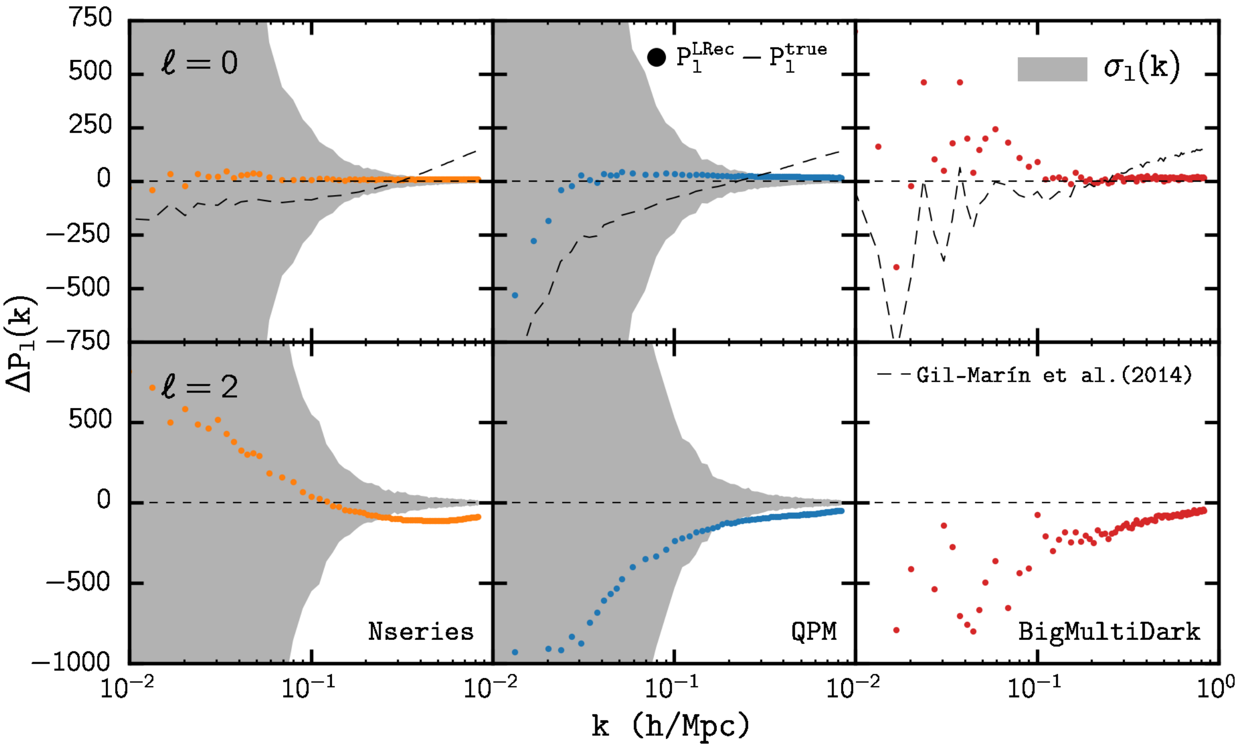

Figure 1.6: The power spectrum residual of the line-of-sight reconstruction (LRec) method (Section 1.4.2), $\Delta P_\ell \equiv P_l^{\mathrm{LRec}} - P_l^{\mathrm{true}}$, for the monopole (top) and quadrupole (bottom) power spectra of the Nseries (left), QPM (middle), and BigMultiDark (right) mock catalogs. We again plot the Nseries and QPM sample variances, $\sigma_l(k)$. The residuals for the monopole show good agreement between P_0^{LRec} and P_0^{true} for the entire k range. For the quadrupole, while the LOS Reconstruction method improves the residuals compared to the NN method at small scales ($k > 0.2\ h/\mathrm{Mpc}$), the residuals remain comparable to sample variance at $k = 0.2\ h/\mathrm{Mpc}$. In the top panels, we include the residuals from the fiber collision correction method of Gil-Marín et al. (2014) (dashed). As the Gil-Marín et al. (2014) method supplements the NN method with adjustments to the constant shot noise term of the estimator, it fails to correct for the k dependence of the effect and is insufficient in accounting for fiber collisions at small scales. We do not include the correction method of Beutler et al. (2014b) because they marginalize over a constant stochasticity term in their analysis so the effect of their correction on $P(k)$ is not straightforward.

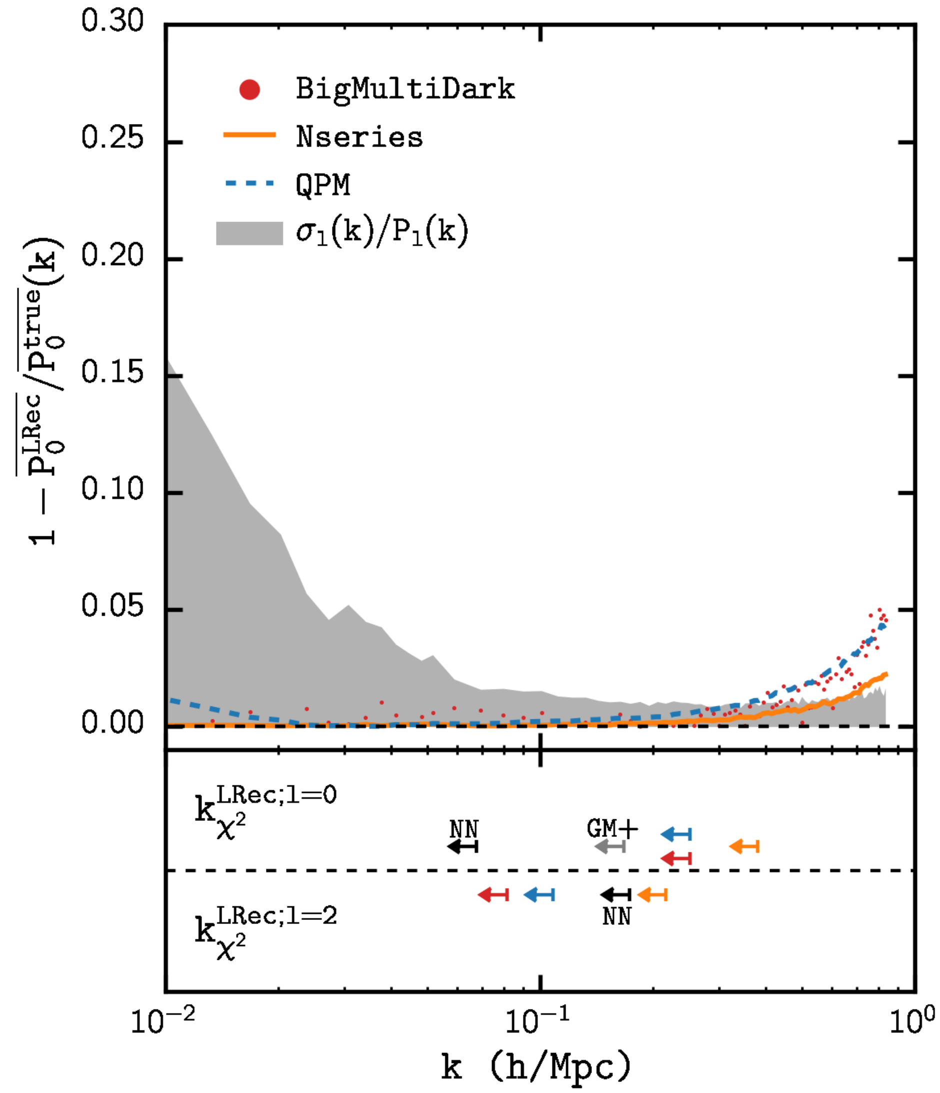

Figure 1.7: *Top Panel*: The normalized residual, $1 - P_l^{\rm LRec}/P_l^{\rm true}$, for the Nseries (orange), QPM (blue), and BigMultiDark (red) monopole power spectra. The normalized sample variance $\sigma_l/P_l(k)$ (gray shaded region) of the Nseries mocks is plotted for comparison. At $k = 0.1$ h/Mpc, where the NN method residuals exceeds sample variance, the average normalized residual for the LRec method is 0.25% compared to 1.5% normalized sample variance. In fact, the average residual stays below the sample variance until $k = 0.53$ h/Mpc. *Bottom Panel*: We mark $k_{\chi^2}^{\rm LRec}$ for the monopole (arrows above the dashed line) and quadrupole (arrows below the dashed line). The average $k_{\chi^2}^{\rm LRec}$ for the mock catalogs are 0.29 and 0.14 h/Mpc for the monopole and quadrupole respectively. For comparison, we mark $k_{\chi^2}^{\rm NN}$ (black) from Section 1.4.1. We also include k_{χ^2} of the Gil-Marín et al. (2014) correction method (gray) for the monopole. The LOS reconstruction method significantly extends k_{χ^2} beyond that of the NN method and Gil-Marín et al. (2014) for $l = 0$. However, it does not improve k_{χ^2} for the quadrupole.

weights, consist of the "nearest-neighbor" galaxy with $w_{\rm fc} > 1$ and the "collided" galaxy with $w_{\rm fc} = 0$. We discard the collided galaxy since the redshifts of collided galaxies are not known in actual observations.

Next for each of the nearest-neighbor galaxies in peak-assigned pairs, we place a new galaxy with $w_{\rm fc} = 1$ at a displacement $d_{\rm peak}$ away from it along the line-of-sight but at the same angular position. The $d_{\rm peak}$ value is sampled from a Gaussian with best-fit $\sigma_{\rm LOS}$ from Table 1.1. The $w_{\rm fc}$ of the "nearest-neighbor" galaxy is then reduced by 1. This process is repeated, in the cases of triplets or higher with $w_{\rm fc} > 2$, until all the nearest-neighbor galaxy in peak-assigned pairs have $w_{\rm fc} = 1$. The resulting *total* catalog will have fewer galaxies with $w_{\rm fc} > 1$ compared to the initial fiber collided catalog. However, the total statistical weight ($\sum_{\rm gal} w_{\rm tot}$) of the catalog, being equal to the total number of galaxies before the collisions are applied, is conserved.

Now that we have the "LOS reconstructed" mock catalog, we measure its power spectrum monopole and quadrupole ($P_l^{\rm LRec}$). In Figure 1.6 we present the power spectrum residual, ($P_l^{\rm LRec} - P_l^{\rm true}$), for $l = 0$ and 2 of the LOS Reconstruction method power spectrum averaged over all the available realizations. We again include the Nseries and QPM sample variance, $\sigma_l(k)$ (grey shaded region) for comparison. In Figure 1.7, we normalize both the residuals and the sample variance by $P_0^{\rm true}$ to better compare $P_0^{\rm LRec}$ and $P_0^{\rm true}$ at different scales and to highlight the small scales.

For the monopole, at the scale where $P_0^{\rm NN}$ deviates from $P_0^{\rm true}$ by more than the sample variance ($k \sim 0.1$ h/Mpc), Figure 1.6 shows that the LOS reconstructed residual is well within the sample variance, $P_0^{\rm LRec} - P_0^{\rm true} < 0.17\,\sigma_0$. Even at the smallest scales measured for our monopole measurements ($k = 0.83$ h/Mpc), well beyond the scales that can be predicted from current models based on perturbation theory, the normalized residuals for the LOS reconstructed method remains at 3.7%. At $k \sim 0.2$ h/Mpc, the average normalized residual is 0.19% compared to the 0.9% normalized sample variance. When we calculate the k_{χ^2} of the LOS reconstruction method for the three mock catalogs, as we did for the NN method in Section 1.4.1, we get the average $k_{\chi^2}^{\mathrm{LRec}} = 0.29\ h/\mathrm{Mpc}$ for the monopole. For each of the mocks, we mark $k_{\chi^2}^{\mathrm{LRec};\, l=0}$ in the lower panel of Figure 1.7 above the dashed horizontal line.

For the monopole, we also include residuals from the fiber collision correction method of Gil-Marín et al. (2014) (dashed) in Figure 1.6. Gil-Marín et al. (2014) corrects for fiber collisions by adjusting the constant shot noise term in the estimator in addition to the NN method (Section 1.4.2.2). However, as the NN method power spectrum residuals reveal in Figure 1.3, the effect is k dependent, especially at $k > 0.1\ h/\mathrm{Mpc}$. So while this correction can reduce the residuals to within sample variance on large scales, it fails to account for the k dependence, which quickly goes on to dominate sample variance at smaller scales, $k > 0.1\ h/\mathrm{Mpc}$.

We also calculate k_{χ^2} for the Gil-Marín et al. (2014) correction method using the mock catalogs, $k_{\chi^2}^{\mathrm{GM+}} = 0.17\ h/\mathrm{Mpc}$ (gray arrow; Figure 1.7), which is significantly lower than that of the LOS Reconstruction method. The LOS reconstruction method better accounts for fiber collisions at all scales. Furthermore, as already discussed, the Gil-Marín et al. (2014) method does *not* provide corrections for the power spectrum quadrupole or higher multipoles, thus Figure 1.3 still applies for $\ell = 2$.

We note that the correction method of Beutler et al. (2014b) is not included in Figure 1.6. Instead of using a fixed value for the constant shot noise as Gil-Marín et al. (2014) does, Beutler et al. (2014b) includes a constant 'stochasticity term', N, in their analysis (see Eq. 40 of Beutler et al. 2014b). This N is within the exponential factor that models the Finger-of-God effect, so their correction is k dependent and impacts the multipoles beyond the monopole. However because Beutler et al. (2014b) marginalizes over N, the effect of this correction is not straightforward. When we use the best-fit parameter values from Beutler et al. (2014b), we find that the correction actually *increases* the effect of fiber collisions on both the monopole and quadrupole. This however, neglects the impact of stochastic bias in the $P(k)$ model. Nevertheless, we also find that no value of N in the Beutler et al. (2014b) correction can simultaneously account for the effect of fiber collisions in both the monopole and quadrupole.

From Figure 1.6 we see that for the quadrupole, the LOS reconstruction method does not sufficiently improve corrections for fiber collisions compared to the NN method. The residuals for $k > 0.2$ h/Mpc are improved compared to Figure 1.3; however, they still exceed the sample variance. Unfortunately, these improvements on small scales come at the cost of increased residuals on large scales. In the k_{χ^2} marked in Figure 1.7 (below the dashed line), we see that the increased residuals at large scales actually make the average $k_{\chi^2}^{\mathrm{NN}} > k_{\chi^2}^{\mathrm{LRec}} = 0.14$ h/Mpc for the quadrupole, although there is significant dispersion between the different simulations with Nseries showing improvements when compared to the NN method while the other two showing worse performance. Consequently, neither the LOS reconstruction method nor the NN method sufficiently account for fiber collisions in the power spectrum quadrupole.

The shortcomings of the LOS reconstruction method for the quadrupole compared to the monopole does not come as a surprise since the quadrupole is more sensitive to getting the correct LOS displacements galaxy by galaxy (not just statistically), as these modify the fingers-of-god effect. In order to make further progress with this method one would have to determine for each galaxy the most likely halo in which it lives (this could be nearby or a distant, chance alignment), determine its velocity dispersion and then assign a LOS displacement consistent with the dispersion and the observed LOS distribution.

Let us now discuss a few attempts that we have implemented along these lines. The first is incorporating more information about the fiber collided pairs in order to better classify correlated and chance alignment pairs. For example, information about larger scale galaxy environment in the form of the N^{th} nearest neighbor distance (d_{nNN}), can be included to parameterize the σ_{LOS} and f_{peak} (Table 1.1) as a function of d_{nNN}. The d_{nNN} in this case is the distance of the n^{th} nearest neighbor of the nearest-neighbor galaxy within the fiber collided pair. Another way the LOS reconstructed method can be improved is by utilizing the photometric redshifts of the collided galaxies to improve the correlated/change alignment pair classification.

We explored the LOS reconstructed method with both of these improvements on the mock catalogs. We find that there is indeed a significant correlation between d_{nNN} and the parameters σ_{LOS} and f_{peak}, which can be exploited. Also, photometric redshifts assigned to collided galaxies based on the $|z_{\text{spec}} - z_{\text{photo}}|/(1 + z_{\text{spec}})$ of actual BOSS photometric redshift catalogs improves classification of correlated versus chance alignment fiber collided pairs, as well. These improvements bring the normalized residuals of the monopole to $\sim 1\%$ at $k = 0.83$ h/Mpc. However, the improvement in the fiber collision correction for the quadrupole is marginal; the effect of fiber collisions at $k = 0.2$ h/Mpc is still comparable to the sample variance. So even with these improvements the LOS reconstructed method is insufficient.

Furthermore, for the Nseries mocks, we find that if we use the LOS reconstructed method with perfectly classified correlated and chance alignment pairs, the residual is roughly half the sample variance at $k \sim 0.2$ h/Mpc and greater than sample variance at $k > 0.35$ h/Mpc. The displacement of the collided galaxy by d_{LOS} sampled from Eq. (1.13) alone causes the power spectrum quadrupole to deviate from the true value at small scales. A method such as the LOS reconstructed method for the quadrupole would require more sophisticated modeling of the fiber collided galaxy pairs that capture the displacements in an object by object basis.

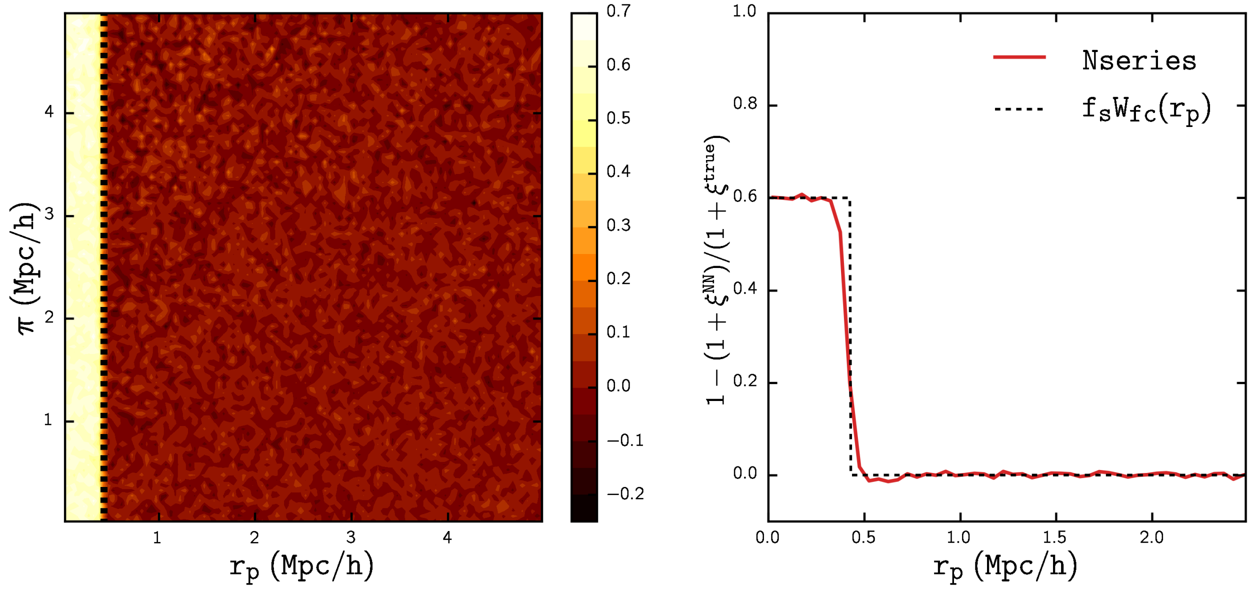

Figure 1.8: $1-(1+\xi^{\mathrm{NN}})/(1+\xi^{\mathrm{true}})$ as a function of transverse displacement, r_p, and line-of-sight displacement π (left). The color bar represents the value of this quantity. Note there is no detectable dependence on π. The dashed vertical line (black) represents the constant $r_p = D_{\mathrm{fc}}(z = 0.55)$ (Section 1.4.3). We also plot $1-(1+\xi^{\mathrm{NN}})/(1+\xi^{\mathrm{true}})$ projected along π (right). In the left panel, the $r_p = D_{\mathrm{fc}}(z = 0.55)$ vertical line and the sharp cut-off of the contour show good agreement with the expected characteristic scale. In the right panel, the projected $1-(1+\xi^{\mathrm{NN}})/(1+\xi^{\mathrm{true}})$ is in good agreement with $f_s W_{\mathrm{fc}}(r_p)$. The agreement in both panels justify the characterization of the effect of fiber collisions on the 2PCF in Eq. (1.21).

As a result of the shortcomings of the LOS reconstructed method for the power spectrum quadrupole, we now present a complementary approach in dealing with fiber collision in power spectrum multipole analyses, which rather than attempting to correct the data before making measurements, computes theoretical predictions of the fiber-collided power spectrum multipoles.

1.4.3 Effective Window Method

The LOS Reconstruction method corrects for fiber collisions in the observed galaxy positions in order to estimate the systematics-free true power spectrum. In power spectrum analyses, this true power spectrum estimate can be compared to model power spectrum for cosmological parameter inference. Alternatively, however, the observed fiber collided power spectrum can be compared to the model power spectrum with the effect of fiber collisions imposed on it. This is the approach we follow from now on.

We proceed as follows. In Section 1.4.3.1 we find that the effect of fiber collisions on the two-point correlation function can be well approximated by a simple analytic expression. Using this, we accurately estimate the effect of fiber collisions on the power spectrum in Fourier space. The effect is a function of the true power spectrum and depends significantly on the power spectrum at small scales, which cannot reliably be modeled from first principles. As a result, in Section 1.4.3.2, we present a practical approach to circumvent this issue and account for the effect of fiber collisions in power spectrum analyses.

1.4.3.1 In Theory

In the BOSS galaxy catalog, which spans the redshifts $0.43 < z < 0.7$, the comoving distance of the 62" fiber collision angular scale ($D_{\rm fc}$) ranges from 0.35 Mpc to 0.52 Mpc. Given the relatively small variation in $D_{\rm fc}$, we assume that throughout the survey redshift the physical scale remains constant as $D_{\rm fc}(z \sim 0.55) = 0.43$Mpc, at the median redshift of the survey. If the physical scale of fiber collisions is constant, fiber collisions will affect the two-dimensional configuration space two-point correlation function, $\xi(r_p, \pi)$, through its effect on galaxy pairs with transverse separations $r_p < D_{\rm fc}$. As no pairs will be found below this characteristic scale, $\xi(r_p, \pi)$ will be -1 for $r_p < D_{\rm fc}$, and note that the same is true for the two-point function in the NN method (since small-r_p pairs are collapsed into zero separation described by weights). On the other hand, at large scales we can approximate $\xi(r_p, \pi)$ by the NN method which preserves the large-scale angular correlation function, thus the effect of fiber collisions on $\xi(r_p, \pi)$ can be analytically characterized by the following relation between the true and the NN two-point functions,

$$\frac{1 + \xi^{\mathrm{NN}}(r_p, \pi)}{1 + \xi^{\mathrm{true}}(r_p, \pi)} = 1 - f_s W_{\mathrm{fc}}(r_p) \tag{1.21}$$

where $W_{\mathrm{fc}}(r_p)$ represents the top-hat function

$$W_{\mathrm{fc}}(r_p) = \begin{cases} 1 & \text{if } r_p < D_{\mathrm{fc}} \\ 0 & \text{otherwise} \end{cases} \tag{1.22}$$

and f_s represents the fraction of the survey area affected by fiber collisions. Note in Eq. (1.21) we have assumed that we can linearly superpose the contributions to the two-point function from regions with and without collisions, and a key property of Eq. (1.21) is that its right hand side does not depend on π, something we test explicitly below. In the BOSS, f_s is precisely known because it corresponds to the fraction of the survey geometry that suffers from fiber collisions. These are the regions that do not have overlapped tiling (Section 1.3). For BOSS DR12 $f_s = 0.6$.

We measure ξ^{NN} and ξ^{true} for the Nseries mock catalogs using the `CUTE` software (Alonso 2012), which uses the standard Landy & Szalay (1993) estimator. ξ^{NN} is calculated from the NN fiber collided Nseries mocks while ξ^{true} is calculated from the Nseries mocks without fiber collisions. Using the measured ξ^{NN} and ξ^{true}, we plot $1 - (1 + \xi^{\mathrm{NN}})/(1 + \xi^{\mathrm{true}})$ averaged over realizations as a function of r_p and π (left) and its projection along π (right) in Figure 1.8. The dashed vertical line (black; left) marking $r_p = D_{\mathrm{fs}}(z = 0.55)$ and $f_s W_{\mathrm{fc}}(r_p)$ (black dashed; right) are plotted for comparison. The agreement between the $\xi(r_p, \pi)$ contours and the $r_p = D_{\rm fc}(z = 0.55)$ cutoff along with the agreement between the projection and $f_s W_{\rm fc}(r_p)$ justify our assumption of a constant physical fiber collision scale. The exact survey tiling of the BOSS sample is imposed on the Nseries mocks, so we expect Figure 1.8 to hold for the BOSS observations. The left panel illustrates the π-independence of the left hand side of Eq. (1.21). The right panel demonstrates that $1 - (1 + \xi^{\rm NN})/(1 + \xi^{\rm true})$ projected along π agrees remarkably well with a top-hat function.

In principle, however, $W_{\rm fc}$ is not necessarily a top-hat function. In fact, in eBOSS, due to the complex targeting scheme involving “knock-outs” from higher priority targeting samples, $W_{\rm fc}$ will not be top-hat function (Zhai et al. in prep). However, these complications are not present in our implementation of collisions; the reason for the deviations from a top-hat function here can be thought as arising from a sum of top-hats of slightly different radii along the line of sight (for fixed angular scale) weighted by the probability of collisions at each depth, leading to a smoother transition than a sharp top-hat function. In principle, our formalism can be improved by including this numerical profile rather than a top-hat, as we shall mention below (see discussion after Eq. 1.34).

With the confirmation of Eq. (1.21), we solve for $\xi^{\rm NN}$:

$$\xi^{\rm NN}(r_p, \pi) \;\; = \;\; \xi^{\rm true}(r_p, \pi) - f_s W_{\rm fc}(r_p)\; (1 + \xi^{\rm true}(r_p, \pi)), \tag{1.23}$$

and to get an expression for the power spectrum, we Fourier transform to get

$$\begin{aligned} \Delta P(\mathbf{k}) \;\; &\equiv \;\; P^{\rm NN}(\mathbf{k}) - P^{\rm true}(\mathbf{k}) \\ &= -f_s\, W_{\rm fc}(\mathbf{k}) - f_s \int \frac{{\rm d}^3 q}{(2\pi)^3} P(\mathbf{q})\, W_{\rm fc}(\mathbf{k} - \mathbf{q}). \end{aligned} \tag{1.24}$$

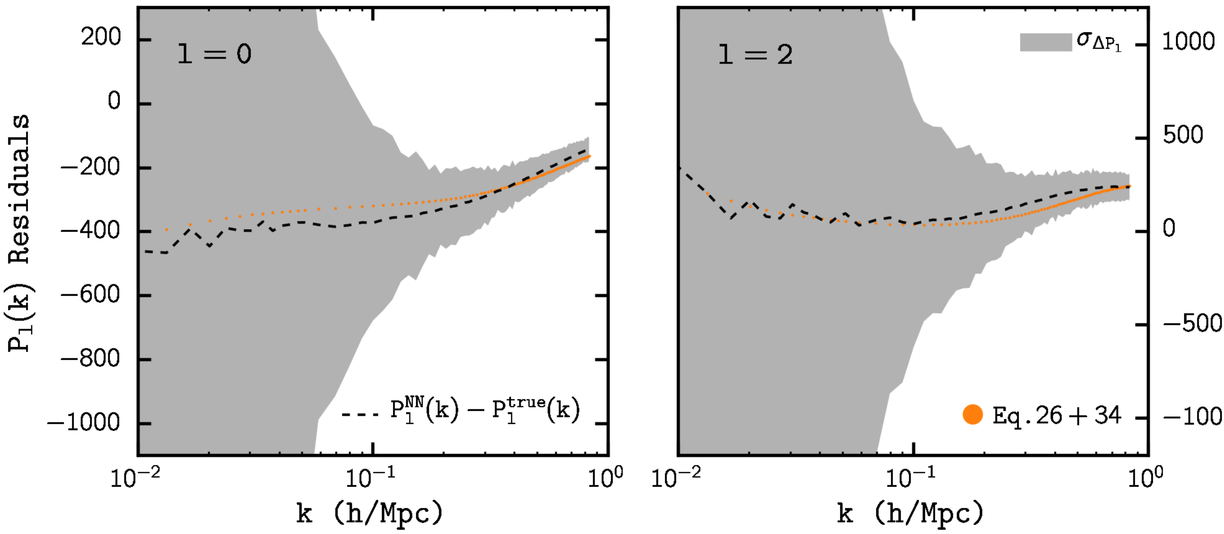

Figure 1.9: Comparison of the power spectrum residuals from NN-corrected fiber collisions $\Delta P_l = P_l^{\rm NN} - P_l^{\rm true}$ (dashed black) with the ΔP_l from the effective window method obtained by adding Eqs. (1.26) and (1.34) (orange) for the monopole (left) and quadrupole (right). The standard deviation of the power spectrum residual, $\sigma_{\Delta P_l}$, for the Nseries mock catalogs is shaded in gray.

We see that the effect of fiber collisions on the true power spectrum can be characterized by two terms: Fourier transform of the top-hat function (corresponding to chance collisions) and the power spectrum convolved with the top-hat function (corresponding to physically correlated pairs). We refer to these two terms as $\Delta P^{\rm uncorr}$ and $\Delta P^{\rm corr}$ respectively. Note that none of these terms is independent of k.

The first term, $\Delta P^{\rm uncorr}$, can be easily obtained:

$$\begin{aligned}\Delta P^{\rm uncorr} &= -f_s\,\widehat{W_{\rm fc}}(\mathbf{k}) = -f_s \int e^{i\mathbf{k}\cdot\mathbf{r}}\,W_{\rm fc}(\mathbf{r})\,d^3r \\ &= -f_s\;2\pi\delta_D(k_{\parallel})\;\pi D_{\rm fc}^2\;W_{\rm 2D}(k_{\perp}D_{\rm fc}). \qquad (1.25)\end{aligned}$$

where $W_{\rm 2D}(x) \equiv 2J_1(x)/x$ is the top-hat function in 2D (a cylinder), and J_1 is a Bessel function of the first kind and of order 1. The multipole contributions of Eq. (1.25) are then

$$\Delta P_l^{\rm uncorr}(k) \;\; = \;\; -f_s\;(2l+1)\mathcal{L}_l(0)\,\frac{(\pi D_{\rm fc})^2}{k}\,W_{\rm 2D}(kD_{\rm fc}), \tag{1.26}$$

where $\mathcal{L}_l$ are the Legendre polynomials. The k^{-1} prefactor here, arising from the delta function in Eq. (1.25) is an approximation for scales smaller than the survey size, since the delta function follows from assuming we can integrate up to infinity along the line of sight in Eq. (1.25). Equation (1.26) gives a correction that alternates in sign as a function of multipole l. Note that since for practical purposes $kD_{\rm fc} \ll 1$, we can expand

$$\begin{aligned}\Delta P_l^{\rm uncorr}(k) \;\; &= \;\; -f_s\pi D_{\rm fc}^2\left(\frac{2\pi}{k}\right)\,\frac{(2l+1)}{2}\,\mathcal{L}_l(0)\\ &\quad\;\; \times\Big(1-\frac{(kD_{\rm fc})^2}{8}+\dots\Big),\end{aligned} \tag{1.27}$$

and for scales involved in typical analysis the first term suffices, which means that the uncorrelated piece of fiber collisions decays as k^{-1} across the relevant range of scales. The magnitude of this uncorrelated effect (chance collisions) is small, given by the effective survey area affected by fiber collisions $f_s\pi D_{\rm fc}^2$ times the wavelength of perturbations $2\pi/k$.

For the correlated piece $\Delta P^{\rm corr}$, we see from Eqs. (1.24) and (1.25) that we need $W_{\rm 2D}(|\mathbf{k}_\perp - \mathbf{q}_\perp|D_{\rm fc})$ for which we can use the addition theorem for 2D top-hat functions (Bernardeau et al., 2002),

$$\begin{aligned}W_{\rm 2D}(|\mathbf{k}_\perp-\mathbf{q}_\perp|D_{\rm fc}) \;\; &= \;\; \sum_{k=0}(k+1)\,U_k(\hat{k}_\perp\cdot\hat{q}_\perp)\\ &\quad\;\; W_{\rm 2D}^{(k/2)}(k_\perp D_{\rm fc})\,W_{\rm 2D}^{(k/2)}(q_\perp D_{\rm fc})\end{aligned} \tag{1.28}$$

where the U_k's are the Chebyshev polynomials and $W_{2\mathrm{D}}^{(k/2)}(x) \equiv 2J_{k+1}(x)/x$. Now, again, as we are interested in scales for which $kD_{\mathrm{fc}} \ll 1$ is an excellent approximation, dropping $\mathcal{O}(k_\perp D_{\mathrm{fc}})^2$ we can just use the $k = 0$ term in this expression. This gives us $W_{2\mathrm{D}}(|\mathbf{k}_\perp - \mathbf{q}_\perp|D_{\mathrm{fc}}) \approx W_{2\mathrm{D}}(q_\perp D_{\mathrm{fc}})$ as expected and leads to,

$$\Delta P^{\mathrm{corr}}(\mathbf{k}) \;\approx\; -f_s \pi D_{\mathrm{fc}}^2 \int \frac{d^2 q_\perp}{(2\pi)^2}\, P(k_\parallel, q_\perp)\, W_{2\mathrm{D}}(q_\perp D_{\mathrm{fc}}) \tag{1.29}$$

This is a simple result, showing that the correlated effect of fiber collisions is proportional to the effective survey area affected by fiber collisions and to the integral of the power spectrum over 2D modes perpendicular to the line of sight smoothed at the fiber collision scale. The multipole components of Eq. (1.29) are, after expanding $P(k_\parallel, q_\perp)$ in multipoles,

$$\Delta P_l^{\mathrm{corr}}(k) \approx -\frac{f_s D_{\mathrm{fc}}^2}{2} \sum_{l'=0}^{\infty} \int_0^\infty q dq P_{l'}(q)\, f_{ll'}(k,q), \tag{1.30}$$

where, again neglecting $\mathcal{O}(kD_{\mathrm{fc}})^2$,

$$\begin{aligned} f_{ll'}(k,q) \;\equiv\; & \left(\frac{2l+1}{2}\right) \int_{\max(-1,-q/k)}^{\min(1,q/k)} d\mu\, \mathcal{L}_l(\mu)\, \mathcal{L}_{l'}(k\mu/q) \\ & \times W_{2\mathrm{D}}(q\, D_{\mathrm{fc}}) \end{aligned} \tag{1.31}$$

This has a simple expression for $l = l'$,

$$f_{ll}(k,q) = f_*(k,q)\, W_{2\mathrm{D}}(q\, D_{\mathrm{fc}}) \left(\frac{k_<}{k_>}\right)^l \tag{1.32}$$

where $f_*(k,q) = q/k$ for $q \leq k$ and unity otherwise, and $k_> = \max(k,q)$ and $k_< = \min(k,q)$.

On the other hand, off the diagonal we have ($l \neq l'$)

$$f_{ll'}(k,q) = f_*(k,q)\, W_{\rm 2D}(q\, D_{\rm fc}) \Big(\frac{2l+1}{2}\Big)\, H_{l_> l_<}\Big(\frac{k_<}{k_>}\Big), \tag{1.33}$$

where $l_> = \max(l, l')$ and similarly $l_<$, and $H_{l_> l_<}(x)$ is a polynomial of degree $l_>$ which vanishes unless l and k are both larger or smaller than l' and q respectively. The first few polynomials are listed in the Appendix 4.8. Since $f_{l>l'}(k<q) = f_{l<l'}(k>q) = 0$ it is convenient to split the integrals depending on whether q is larger or smaller than k, which leads to

$$\begin{aligned} \Delta P_l^{\rm corr}(k) \ \approx\ & -\frac{f_s D_{\rm fc}^2}{2}\Bigg[\sum_{l'\leq l}\int_0^k q dq\, P_{l'}(q)\, f_{ll'}(q\leq k) \\ & + \sum_{l'\geq l}\int_k^\infty q dq\, P_{l'}(q)\, f_{ll'}(q\geq k)\Bigg], \end{aligned} \tag{1.34}$$

which shows that the change of power spectrum multipole l due to correlated fiber collisions comes from long modes of lower multipoles ($l' \leq l$) and short modes of higher multipoles ($l' \geq l$). Going back to the results displayed in Figure 1.8, we can now formulate how our results change if we use the observed numerical profile in the right panel of Figure 1.8 (red line) instead of the top-hat (black dashed). One can check that to leading order in $kD_{\rm fc}$, which is all we are using in this paper, our expression for the uncorrelated and correlated change in power are valid as long as we replace the 2D top-hat by the numerical profile in Eq. (1.31), and redefine the scale $D_{\rm fc}$ that appears in Eqs. (1.27) and (1.30) from the area of the numerical profile, that is

$$\int d^2 r_\perp\, W_{\rm 2D}(\mathbf{r}_\perp) \equiv \pi\, D_{\rm fc}^2 \tag{1.35}$$

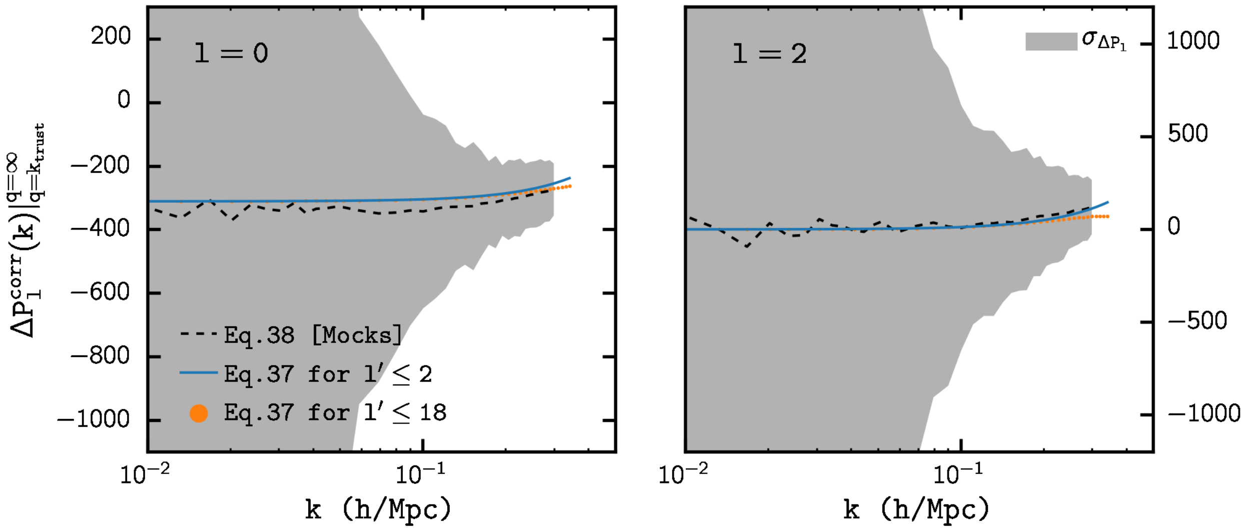

Figure 1.10: Comparison of the correlated power spectrum residuals from unreliable modes obtained from mocks (dashed), Eq. (1.38), to the polynomial approximation of Eq. (1.37) for $l' \leq 18$ (orange). The left and right panels correspond to $l = 0$ and 2 respectively. The gray shaded region is the standard deviation for the Nseries ($P_l^{\rm NN} - P_l^{\rm true}$). We also include Eq. (1.37) evaluated only for $l' \leq 2$ (blue). The agreement between Eq. (1.37) for $l' \leq 2$ and Eq. (1.38) demonstrate that while higher orders of l' are necessary to properly model $\Delta P_l^{(}k)$ at higher k values, for $k < k_{\rm trust}$ (0.3 h/Mpc above) $l' \leq 2$ are sufficient.

We now proceed to testing these results, for which we need the true power spectrum multipoles down to small scales to feed into Eq. (1.34). Unfortunately, in the nonlinear regime the multipole expansion is not very efficient (in the sense that the amplitude of multipoles does not decrease sharply with increasing multipole), so a large number of multipoles l' is required to capture the contribution from small scale modes. Measuring multipoles higher than the hexadecapole for realistic survey geometries using our estimator becomes expensive due to the number of Fast Fourier Transforms (FFTs) that needs to be computed, and even for the most efficient version of the multipole estimators that requires only 7 FFTs one would worry about increased cosmic variance (see discussion in Scoccimarro 2015).

A more efficient approach is to use the Nseries simulation boxes to test Eqs. (1.25) and (1.34). The Nseries simulation boxes are the original simulations where the Nseries mocks were cut out from (Section 1.3). Since the Nseries mocks are cut outs of the boxes, discrepancies in their power spectra are caused by the BOSS survey geometry and occur mainly at the largest scales, $k < 0.05\ h/\mathrm{Mpc}$ (Beutler et al., 2014b; Grieb et al., 2016). At smaller scales, the difference between the power spectrum monopole, quadrupole and hexadecapole of Nseries mocks versus the Nseries boxes are negligible. Therefore, we calculate the $P_{l'}(q)$ from the Nseries simulation box, using periodic boundary conditions, which only requires one FFT and go up to $q = 43.5\ h/\mathrm{Mpc}$ and $l' = 18$ to compute the corrections predicted by Eq. (1.34).

In Figure 1.9, we compare $\Delta P_l = \Delta P_l^{\mathrm{corr}} + \Delta P_l^{\mathrm{uncorr}}$ calculated from the Nseries Box power spectrum multipoles using Eqs. (1.25) and (1.34) (orange) to the Nseries mock catalogs power spectrum residuals, $\Delta P_l = P_l^{\mathrm{NN}} - P_l^{\mathrm{true}}$ (dashed). The left panel compares the monopoles ($l = 0$) while the right panel compares the quadrupoles ($l = 2$). We also include in the gray shaded region, the standard deviation of Nseries mock catalogs power spectrum residuals, $\sigma_{\Delta P_l}$. For both the monopole and quadrupole, the predictions (orange) agree with the measured residuals from NN-corrected fiber collisions (dashed black) well within the errors throughout the probed k range up to $k = 0.83\ h/\mathrm{Mpc}$. At low-k, the downturn (upturn) in the monopole (quadrupole) is due to the contribution of the k^{-1} uncorrelated piece. The overall quality of the agreement demonstrates that the effective window method can be used to robustly estimate the effect of fiber collisions on $P_l(k)$. Furthermore, with its excellent performance for the quadrupole, the effective window approach provides an improvement over the LOS reconstruction method (Section 1.4.2).

1.4.3.2 In Practice

There are, however, practical limitations to the effective window model as it described above. The $\Delta P_l^{\mathrm{corr}}$ calculations in Eq. (1.34) involves integrating the power spectrum over the q range of 0 to ∞. While this integral converges for $q \approx 10\ h/\mathrm{Mpc}$ for both monopole and quadrupole, in practice one cannot compute reliably the power spectrum multipoles down to these scales. We now discuss a way to overcome this issue.

Let k_{trust} represent the scale up to which we can calculate reliably power spectrum multipoles. We therefore split the second term in Eq. (1.34) into a reliable piece (integration from k to k_{trust}) and an unreliable piece (integration from k_{trust} to ∞), so schematically

$$\Delta P_l^{\mathrm{corr}} = \Delta P_l^{\mathrm{corr}} \Bigg|_{q=0}^{q=k_{\mathrm{trust}}} + \Delta P_l^{\mathrm{corr}} \Bigg|_{q=k_{\mathrm{trust}}}^{q=\infty}. \tag{1.36}$$

The first term can be reliably calculated from first principles since it involves modes from $q = 0$ to $q = k_{\mathrm{trust}}$ and corresponds to the first term plus the reliable piece of the second term in Eq. (1.34). Now, the key fact is that because the second term in Eq. (1.36) only depends on k through $f_{ll'}(q \geq k)$, from Eqs. (1.32-1.33) it follows that the k-dependence of the unreliable term is simply a polynomial in k,

$$\Delta P_l^{\text{corr}}\Big|_{q=k_{\text{trust}}}^{q=\infty} = \sum_{n=0,2,4...} C_{l,n}\, k^n. \tag{1.37}$$

The coefficients of the polynomial, $C_{l,n}$, are obtained by collecting powers of k from the sum over the H-polynomial contributions to the second term in Eq. (1.34). How important are these unreliable contributions? In order to test this, in Figure 1.10 we calculate $C_{l,n}$ from the $P_{l'}(q)$ multipoles measured from the Nseries simulation boxes (blue and orange for terms up to $l' = 2$ and 18 respectively) and compare to (black dashed)

$$\Delta P_l^{\text{Nseries}}(k) - \Delta P_l^{\text{uncorr}}(k) - \Delta P_l^{\text{corr}}(k)\Big|_{q=0}^{q=k_{\text{trust}}} \tag{1.38}$$

where $\Delta P_l^{\text{Nseries}}$ is the power spectrum residual $P_l^{\text{NN}} - P_l^{\text{true}}$ for the Nseries mocks (Figure 1.9). We once again include the standard deviation of the power spectrum residual in shaded gray. The agreement between Eq. (1.37) and Eq. (1.38) is more or less equivalent to the agreement seen in Figure 1.9, which includes uncorrelated and reliable correlated contributions as well; this should of course not come as a surprise.

More importantly, when we examine the contribution to $\Delta P_l^{\text{corr}}|_{k_{\text{trust}}}^{\infty}$ from each individual l' order term of the Eq. (1.37) polynomial, we find that the main contributors at $k < k_{\text{trust}} \sim 0.3\ h/\text{Mpc}$ are the $l' \leq 2$ order terms. In fact, the higher order ($l' > 2$) terms of the polynomial contribute at higher k. For instance, the $l' = 4, 6$, and 8 terms only begin to significantly contribute at scales of $k > 0.3,\ 0.45$, and $0.6\ h/\text{Mpc}$ respectively, which is not surprising since higher k powers come together with increasing inverse powers of q and thus suppress the value of the coefficients that result from integrating over small-scale modes. Hence, when we plot Eq. (1.37) for just $l' \leq 2$ (blue) in Figure 1.10, we find that it is in good agreement with both Eq. (1.37) for $l' \leq 18$ and Eq. (1.38). We also note that for $l = 2$,

$C_{2,l'=0} = 0$ so the main contribution to $\Delta P_2^{\text{corr}}(k < k_{\text{trust}})|_{k_{\text{trust}}}^{\infty}$ comes solely from the $l' = 2$ term of the polynomial.

To use the effective window method for cosmological inference, we can utilize the fact that Eq. (1.37) with only $l' \leq 2$ terms provides an accurate estimate of the unreliable correlated change in power (Figure 1.10). In cosmological analyses, the coefficients $C_{l,0}$ and $C_{l,2}$ can be nuisance parameters with priors obtained from mock catalogs. More specifically, for the quadrupole, since $C_{2,0} = 0$ only one nuisance parameter is necessary. Meanwhile for the monopole, a constant shot noise term is typically already included as a nuisance parameter in the analysis (Beutler et al. 2014b, 2016; Grieb et al. 2016; Gil-Marín et al. 2016b) so there is also only one extra nuisance parameter for $l = 0$. Therefore, by adding $C_{l,2}$ as nuisance parameters to cosmological inference analyses of the power spectrum multipoles, we can use the effective window method to robustly marginalize over the effects of fiber collision for the entire k range of power spectrum models based on perturbation theory.

1.5 Summary and Conclusions

Using simulated mock catalogs designed specifically for interpreting BOSS clustering measurements with realistically imposed fiber collisions, we demonstrate that the Nearest Neighbor method (NN), most common used for dealing with fiber collisions, is insufficient in accounting for the effect of fiber collisions on the galaxy power spectrum monopole and quadrupole. Although fiber collisions have little significant effect on the power spectrum at large scales, their effect quickly overtakes sample variance on scales smaller than $k \approx 0.1$ h/Mpc. At $k \sim 0.3$ h/Mpc fiber collisions have over a 7.3% and 73% impact on the power spectrum monopole and quadrupole, respectively. The effect is equivalent to 7.3 and 2.5 times the sample variance of CMASS for $\delta k \approx 0.01$ h/Mpc, leading to a binning-independent scale of validity of the NN method of $k_{\chi^2} = 0.068$ h/Mpc for the monopole and $k_{\chi^2} = 0.17$ h/Mpc for the quadrupole (see bottom panel of Figure 1.7). Consequently at these scales, measurements of the power spectrum becomes dominated by the systematic effects of fiber collisions.

Some recent methods (Beutler et al. 2014b; Gil-Marín et al. 2014; Beutler et al. 2016; Grieb et al. 2016; Gil-Marín et al. 2016b) have supplemented the NN method with adjustments to the constant shot noise term in the power spectrum estimator. While these methods improve the overall residual for the monopole, e.g. $k_{\chi^2} = 0.17$ h/Mpc for the method by Gil-Marín et al. (2014), they fail to account for the k-dependence of the systematic effect on smaller scales. Furthermore, since the quadrupole does not have a shot noise term, these methods provide no improvements for $l \geq 2$.

In this paper, we first model the distribution of the line-of-sight displacement between fiber collided pairs using mock catalogs. From the model, we statistically reconstruct the clustering of fiber collided galaxies that reside in the same halo. This, combined with the actual shot noise subtraction of the power spectrum estimator that accounts for chance alignments, leads to our LOS Reconstruction method that recovers very well the true power spectrum monopole from fiber collided data. As an added advantage, the method only relies on parameters (σ_{LOS} and f_{peak}) measured from the actual observations. This makes the performance of the method independent from the accuracy of the mock catalogs, which are known to be unreliable at small scales.

Using the LOS Reconstruction method, we can recover the true power spectrum monopole to scales well beyond previous methods. The LOS Reconstruction monopole power spectrum residuals remain within sample variance until $k \sim 0.53$ h/Mpc and k_{χ^2} extends to 0.29 h/Mpc. However, for the power spectrum quadrupole at $k = 0.2$ h/Mpc, the LOS Reconstruction method only reduces the discrepancy between the fiber collided $P_2(k)$ and the true $P_2(k)$ to roughly the sample variance. Therefore, the true monopole power spectrum estimate from the LOS reconstruction method can be compared to the systematics free predicted power spectrum monopole to infer the cosmological parameters of interest without biases from fiber collisions, but for the quadrupole power spectrum the method is not a substantial improvement over previous methods. We trace this problem to the fact that the quadrupole is more sensitive to the object by object finger of god effect, while the LOS reconstruction works only statistically starting from the distribution of close pairs.

To improve on the LOS reconstruction results we develop the effective window method which, rather than attempting to correct the data before making measurements, computes theoretical predictions of the fiber-collided power spectrum multipoles. In this approach, we approximate the effect that fiber collisions have on the two-dimensional configuration space two-point correlation function of the NN method as a scaled top-hat function. Then the effect of fiber collisions can be written as the sum of two contributions: 1) that of uncorrelated chance collisions, with an amplitude proportional to the the effective survey area affected by fiber collisions times the wavelength of perturbations, and 2) that of correlated collisions, which is also proportional to the effective survey area affected by fiber collisions and to the integral of the power spectrum over 2D modes perpendicular to the line of sight smoothed at the fiber collision scale.

Using high resolution mock catalogs, we demonstrate that our analytic prescription accurately models the power spectrum residuals from the NN method to within sample variance of BOSS volumes at $k < 0.83\ h/\mathrm{Mpc}$ for both the monopole and quadrupole when the true power spectrum is known down to small scales from simulations, allowing to compute the fiber-collided predictions. Since typically we do not have fast reliable ways of computing the small scale power spectrum, we develop a practical approach when the power spectrum predictions are reliable up to some scale k_{trust}. We split the contributions of the correlated fiber collisions effect into a piece that can be calculated reliably as it depends on large-scale modes, and an unreliable piece that depends on modes that are not under control. We show that the latter piece can be written as polynomials in k, and demonstrate that for scales up to $k \sim 0.3\ h/\mathrm{Mpc}$, the unreliable contribution can be accurately estimated by a quadratic polynomial in k. In principle, this method can be applied to larger k_{trust} than used here as a reasonable example ($k_{\mathrm{trust}} = 0.3\ h/\mathrm{Mpc}$).

Therefore, using the effective window method we can model the fiber collided power spectrum as the systematics-free power spectrum plus three contributions due to fiber collisions: an uncorrelated piece (independent of the model power spectrum), a calculable piece (which involves integrating the model power spectrum over 2D long-wavelength modes perpendicular to the line of sight), and an unreliable contribution that is a quadratic polynomial, $C_{l,0} + C_{l,2}\, k^2$. While the precise values of $C_{l,n}$ cannot be robustly predicted in practice because of its dependence on small scale power, the coefficients can be treated as nuisance parameters in the analysis. Typically a constant shot noise term is already included as a nuisance parameter, while the constant contribution vanishes for higher multipoles, therefore only one extra parameter per multipole is required (the k^2 corrections). For cosmological parameter inference, the fiber collided model power spectrum can be compared directly to the observed fiber collided power spectrum. Then by marginalizing over these free coefficients, we marginalize over the effect of small-scale power induced fiber collisions on the power spectrum, which allows us to robustly infer the cosmological parameters of interest.

The fiber collision correction methods we present will enable us to robustly account for the effects of fiber collisions in galaxy clustering analyses to the smallest scales allowed by theoretical predictions. They can also be extended to future surveys such as eBOSS or any other large fiber-fed surveys that suffer from systematic effects of fiber collisions. Our fiber collision correction method can also be extended to higher order clustering statistics such as bispectrum (Hahn et al., in prep.). We will use the methods presented in this paper to analyze the galaxy power spectrum and bispectrum multipoles in future work.

Acknowledgements

CHH and MRB were supported by NSF-AST-1109432 and NSF-AST-1211644. SRT is grateful for support from the Campus de Excelencia Internacional UAM/CSIC. We thank A. I. Malz, Mohammadjavad Vakili, Johan Comparat and particularly David W. Hogg for helpful discussions. CHH also thanks the Instituto de Física Teoórica (UAM/CSIC) and particularly Francisco Prada for their hospitality during his summer visit, where part of this work was completed.

Chapter 2

Approximate Bayesian Computation in Large Scale Structure: constraining the galaxy-halo connection

This Chapter is joint work with Mohammadjavad Vakili (NYU), Kilian Walsh (NYU), Andrew P. Hearin (Yale), David W. Hogg (NYU), and Duncan Campbell (Yale) submitted to the *Monthly Notices of the Royal Astronomical Society* as Hahn et al. (2017).

2.1 Chapter Abstract

Standard approaches to Bayesian parameter inference in large scale structure assume a Gaussian functional form (chi-squared form) for the likelihood. This assumption, in detail, cannot be correct. Likelihood free inferences such as Approximate Bayesian Computation (ABC) relax these restrictions and make inference possible without making any assumptions on the likelihood. Instead it relies on a forward generative model of the data and a metric for measuring the distance between the model and data. In this work, we demonstrate that ABC is feasible for LSS parameter inference by using it to constrain parameters of the halo occupation distribution (HOD) model for populating dark matter halos with galaxies.

Using specific implementation of ABC supplemented with Population Monte Carlo importance sampling, a generative forward model using HOD, and a distance metric based on galaxy number density, two-point correlation function, and galaxy group multiplicity function, we constrain the HOD parameters of mock observation generated from selected "true" HOD parameters. The parameter constraints we obtain from ABC are consistent with the "true" HOD parameters, demonstrating that ABC can be reliably used for parameter inference in LSS. Furthermore, we compare our ABC constraints to constraints we obtain using a pseudo-likelihood function of Gaussian form with MCMC and find consistent HOD parameter constraints. Ultimately our results suggest that ABC can and should be applied in parameter inference for LSS analyses.

2.2 Introduction

Cosmology was revolutionized in the 1990s with the introduction of likelihoods—probabilities for the data given the theoretical model—for combining data from different surveys and performing principled inferences of the cosmological parameters (White & Scott 1996; Riess et al. 1998). Nowhere has this been more true than in cosmic microwave background (CMB) studies, where it is nearly possible to analytically evaluate a likelihood function that involves no (or minimal) approximations (Oh et al. 1999, Wandelt et al. 2004, Eriksen et al. 2004, Planck Collaboration et al. 2014b, 2015a).

Fundamentally, the tractability of likelihood functions in cosmology flows from the fact that the initial conditions are exceedingly close to Gaussian in form (Planck Collaboration et al. 2015b,c), and that many sources of measurement noise are also Gaussian (Knox 1995; Leach et al. 2008). Likelihood functions are easier to write down and evaluate when things are closer to Gaussian, so at large scales and in the early universe. Hence likelihood analyses are ideally suitable for CMB data.

In large-scale structure (LSS) with galaxies, quasars, and quasar absorption systems as tracers, formed through nonlinear gravitational evolution and biasing, the likelihood *cannot* be Gaussian. Even if the initial conditions are perfectly Gaussian, the growth of structure creates non-linearities which are non-Gaussian (see Bernardeau et al. 2002 for a comprehensive review). Galaxies form within the density field in some complex manner that is modeled only effectively (Dressler 1980; Kaiser 1984; Santiago & Strauss 1992; Steidel et al. 1998; see Somerville & Davé 2015 for a recent review). Even if the galaxies were a Poisson sampling of the density field, which they are not (Mo & White 1996; Somerville et al. 2001; Casas-Miranda et al. 2002), it would be tremendously difficult to write down even an approximate likelihood function (Ata et al. 2015).

The standard approach makes the strong assumption that the likelihood function for the data can be approximated by a pseudo-likelihood function that is a Gaussian probability density in the space of the two-point correlation function estimate. It is also typically limited to (density and) two-point correlation function (2PCF) measurements, assuming that these measurements constitute sufficient statistics for the cosmological parameters. As Hogg (in preparation) demonstrates, the assumption of a Gaussian pseudo-likelihood function cannot be correct (in detail) at any scale, since a correlation function, being related to the variance of a continuous field, must satisfy non-trivial positive-definiteness requirements. These requirements truncate function space such that the likelihood in that function space could never be Gaussian. The failure of this assumption becomes more relevant as the correlation function becomes better measured, so it is particularly critical on intermediate scales, where neither shot noise nor cosmic variance significantly influence the measurement.

Fortunately, these assumptions are not required for cosmological inferences, because high-precision cosmological simulations can be used to directly calculate LSS observables. Therefore, we can simulate not just the one- or two-point statistics of the galaxies, but also any higher order statistics that might provide additional constraining power on a model. In principle, there is therefore no strict need to rely on these common but specious analysis assumptions as it is possible to calculate a likelihood function directly from simulation outputs.

Of course, any naive approach to sufficiently simulating the data would be ruinously expensive. Fortunately, there are principled, (relatively) efficient methods for minimizing computation and delivering correct posterior inferences, using only a data simulator and some choices about statistics. In the present work, we use Approximate Bayesian Computation—ABC—which provides a *rejection sampling* framework (Pritchard et al. 1999) that relaxes the assumptions of the traditional approach.

ABC approximates the posterior probability distribution function (model given the data) by drawing proposals from the prior over the model parameters, simulating the data from the proposals using a forward generative model, and then rejecting the proposals that are beyond a certain threshold "distance" from the data, based on summary statistics of the data. In practice, ABC is used in conjunction with a more efficient sampling operation like Population Monte Carlo (PMC; Del Moral et al. 2006). PMC initially rejects the proposals from the prior with a relatively large "distance" threshold. In subsequent steps, the threshold is updated adaptively, and samples from the proposals that have passed the previous iteration are subjected to the new, more stringent, threshold criterion (Beaumont et al. 2009). In principle, the distance metric can be any positive definite function that compares various summary statistics between the data and the simulation.

In the context of astronomy, this approach has been used in a wide range of topics including image simulation calibration for wide field surveys (Akeret et al. 2015a), the study of the morphological properties of galaxies at high redshifts (Cameron & Pettitt 2012a), stellar initial mass function modeling (Cisewski et al. in preparation), and cosmological inference with with weak-lensing peak counts (Lin & Kilbinger 2015a; Lin et al. 2016a), Type Ia Supernovae (Weyant et al. 2013a), and galaxy cluster number counts (Ishida et al. 2015a).

In order to demonstrate that ABC can be tractably applied to parameter estimation in contemporary LSS analyses, we narrow our focus to inferring the parameters of a Halo Occupation Distribution (HOD) model. The foundation of HOD predictions is the halo model of LSS, that is, collapsed dark matter halos are biased tracers of the underlying cosmic density field (Press & Schechter 1974; Bond et al. 1991; Cooray & Sheth 2002). The HOD specifies how the dark matter halos are populated with galaxies by modeling the probability that a given halo hosts N galaxies subject to some observational selection criteria (Lemson & Kauffmann 1999; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Zheng et al. 2005a). This statistical prescription for connecting galaxies to halos has been remarkably successful in reproducing the galaxy clustering, galaxy–galaxy lensing, and other observational statistics (Rodríguez-Torres et al. 2015; Miyatake et al. 2015), and is a useful framework for constraining cosmological parameters (van den Bosch et al. 2003; Tinker et al. 2005b; Cacciato et al. 2013; More et al. 2013) as well as galaxy evolution models (Conroy & Wechsler 2009b; Tinker et al. 2011; Leauthaud et al. 2012b; Behroozi et al. 2013a; Tinker et al. 2013, Walsh et al. in preparation).

More specifically, we limit our scope to a likelihood analysis of HOD model parameter space, keeping cosmology fixed. We forward model galaxy survey data by populating pre-built dark matter halo catalogs obtained from high resolution N-body simulations (Klypin et al. 2011; Riebe et al. 2011) using Halotools[1] (Hearin et al. 2016a), an open-source package for modeling the galaxy-halo connection. Equipped with the forward model, we use summary statistics such as number density, two-point correlation function, galaxy group multiplicity function (GMF) to infer HOD parameters using ABC.

In Section 2.3 we discuss the algorithm of the ABC-PMC prescription we use in our analyses. This includes the sampling method itself, the HOD forward model, and the computation of summary statistics. Then in Section 2.4.1, we discuss the mock galaxy catalog, which we treat as observation. With the specific choices of ABC-PMC ingredients, which we describe in Section 2.4.2, in Section 2.4.3 we present the results of our parameter inference using two sets of summary statistics, number density and 2PCF and number density and GMF. We also include in our results, analogous parameter constraints from the standard MCMC approach, which we compare to ABC results in detail, Section 2.4.4. Finally, we discuss and conclude in Section 4.7.

2.3 Methods

2.3.1 Approximate Bayesian Computation

ABC is based on rejection sampling, so we begin this section with a brief overview of rejection sampling. Broadly speaking, rejection sampling is a Monte Carlo method used to draw samples from a probability distribution, $f(\alpha)$, which is difficult to directly sample. The strategy is to draw samples from an instrumental distribution $g(\alpha)$ that satisfies the condition $f(\alpha) < Mg(\alpha)$ for all α, where $M > 1$ is some scalar multiplier. The purpose of the instrumental distribution $g(\alpha)$ is that it is easier to sample than $f(\alpha)$ (see Bishop 2007

[1] http://halotools.readthedocs.org

and refernces therein).

In the context of simulation-based inference, the ultimate goal is to sample from the joint probability of a simulation X and parameters $\vec{\theta}$ given observed data D, the posterior probability distribution. From Bayes rule this posterior distribution can be written as

$$p(\vec{\theta}, X|D) = \frac{p(D|X)p(X|\vec{\theta})\pi(\vec{\theta})}{\mathcal{Z}} \tag{2.1}$$

where $\pi(\vec{\theta})$ is the prior distribution over the parameters of interest and $\mathcal{Z}$ is the evidence,

$$\mathcal{Z} = \int d\vec{\theta}\, dX\ p(D|X)p(X|\vec{\theta})\pi(\vec{\theta}), \tag{2.2}$$

where the domain of the integral is all possible values of X and $\vec{\theta}$. Since $p(\vec{\theta}, X|D)$ cannot be directly sampled, we use rejection sampling with instrumental distribution

$$q(\vec{\theta}, X) = p(X|\vec{\theta})\pi(\vec{\theta}) \tag{2.3}$$

and the choice of

$$M = \frac{\max p(D|X)}{\mathcal{Z}} > 1. \tag{2.4}$$

Note that we do not ever need to know $\mathcal{Z}$. The choices of $q(\vec{\theta}, X)$ and M satisfy the condition

$$p(\vec{\theta}, X|D) < Mq(\vec{\theta}, X) \tag{2.5}$$

so we can sample $p(\vec{\theta}, X|D)$ by drawing $\vec{\theta}, X$ from $q(\vec{\theta}, X)$. In practice, this is done by first drawing $\vec{\theta}$ from the prior $\pi(\vec{\theta})$ and then generating a simulation $X = f(\vec{\theta})$ via the forward model. Then $\vec{\theta}, X$ is accepted if

$$\frac{p(\vec{\theta}, X|D)}{Mq(\vec{\theta}, X)} = \frac{p(D|X)}{\max p(D|X)} > u \tag{2.6}$$

where u is drawn from $\texttt{Uniform}[0, 1]$. By repeating this rejection sampling process, we sample the distribution $p(\vec{\theta}, X|D)$ with the set of $\vec{\theta}$ and X that are accepted.

At this stage, ABC distinguishes itself by postulating that $p(D|X)$, the probability of observing data D given simulation X (*not* the likelihood), is proportional to the probability of the distance between the data and the simulation X being less than an arbitrarily small threshold ϵ

$$p(D|X) \propto p(\rho(D, X) < \epsilon) \tag{2.7}$$

where $\rho(D, X)$ is the distance between the data D and simulation X. Eq. 2.7 along with the rejection sampling acceptance criteria (Eq. 2.6), leads to the acceptance criteria for ABC: $\vec{\theta}$ is accepted if $\rho(D, X) < \epsilon$.

The distance function is a positive definite function that measures the closeness of the data and the simulation. The distance can be a vector with multiple components where each component is a distance between a single summary statistic of the data and that of the simulation. In that case, the threshold ϵ in Eq. 2.7 will also be a vector with the same dimensions. $\vec{\theta}$ is accepted if the distance vector is less than the threshold vector for every component.

The ABC procedure begins, in the same fashion as rejection sampling, by drawing $\vec{\theta}$ from the prior distribution $\pi(\vec{\theta})$. The simulation is generated from $\vec{\theta}$ using the forward model, $X = f(\vec{\theta})$. Then the distance between the data and simulation, $\vec{\rho}(D, X)$, is calculated and compared to $\vec{\epsilon}$. If $\vec{\rho}(D, X) < \vec{\epsilon}$, $\vec{\theta}$ is accepted. This process is repeated until we are left with a sample of $\vec{\theta}$ that all satisfy the distance criteria. This final ensemble approximates the posterior probability distribution $p(\vec{\theta}, X|D)$.

As it is stated, the ABC method poses some practical challenges. If the threshold ϵ is arbitrarily large, the algorithm essentially samples from the prior $\pi(\vec{\theta})$. Therefore a sufficiently small threshold is necessary to sample from the posterior probability distribution. However, an appropriate value for the threshold is not known *a priori*. Yet, even if an appropriate threshold is selected, a small threshold requires the entire process to be repeated for many draws of $\vec{\theta}$ from $\pi(\vec{\theta})$ until a sufficient sample is acquired. This often presents computation challenges.

We overcome some of the challenges posed by the above ABC method by using a Population Monte Carlo (PMC) algorithm as our sampling technique. PMC is an iterative method that performs rejection sampling over a sequence of $\vec{\theta}$ distributions ($\{p_1(\vec{\theta}), ..., p_T(\vec{\theta})\}$ for T iterations), with a distance threshold that decreases at each iteration of the sequence.

As illustrated in Algorithm 1, for the first iteration $t = 1$, we begin with an arbitrarily large distance threshold ϵ_1. We draw $\vec{\theta}$ (hereafter referred to as particles) from the prior distribution $\pi(\vec{\theta})$. We forward model the simulation $X = f(\vec{\theta})$, calculate the distance $\rho(D, X)$, compare this distance to ϵ_1, and then accept or reject the $\vec{\theta}$ draw. Because we set ϵ_1 arbitrarily large, the particles essentially sample the prior distribution. This process is repeated until we accept N particles. We then assign equal weights to the N particles: $w_1^i = 1/N$.

For subsequent iterations ($t > 1$) the distance threshold is set such that $\epsilon_{i,t} < \epsilon_{i,t-1}$ for all components i. Although there is no general prescription, the distance threshold $\epsilon_{i,t}$ can be assigned based on the empirical distribution of the accepted distances of the previous iteration, $t - 1$. In Weyant et al. 2013a, for instance, the threshold of the second iteration is set to the 25th percentile of the distances in the first iterations; afterwards in the subsequent iterations, t, ϵ_t is set to the 50th percentile of the distances in the previous $t - 1$ iteration. Alternatively, Lin & Kilbinger 2015a set ϵ_t to the median of the distances from the previous

Algorithm 1 The procedure for ABC-PMC

```
if $t = 1$ : then
    for $i = 1, ..., N$ do
        // This loop can now be done in parallel for all $i$
        while $\rho(X, D) > \epsilon_t$ do
            $\vec{\theta}_t^* \leftarrow \pi(\vec{\theta})$
            $X = f(\vec{\theta}_t^*)$
        end while
        $\vec{\theta}_t^{(i)} \leftarrow \vec{\theta}_t^*$
        $w_t^{(i)} \leftarrow 1/N$
    end for
end if
if $t = 2, ..., T$ : then
    for $i = 1, ..., N$ do
        // This loop can now be done in parallel for all $i$
        while $\rho(X, D) > \epsilon_t$ do
            Draw $\vec{\theta}_t^*$ from $\{\vec{\theta}_{t-1}\}$ with probabilities $\{w_{t-1}\}$
            $\vec{\theta}_t^* \leftarrow K(\vec{\theta}_t^*, .)$
            $X = f(\vec{\theta}_t^*)$
        end while
        $\vec{\theta}_t^{(i)} \leftarrow \vec{\theta}_t^*$
        $w_t^{(i)} \leftarrow \pi(\vec{\theta}_t^{(i)}) / \big( \sum_{j=1}^{N} w_{t-1}^{(i)} K(\vec{\theta}_{t-1}^{(j)}, \vec{\theta}_t^{(i)}) \big)$
    end for
end if
```

iteration. In Section 2.4, we describe our prescription for the distance threshold, which follows Lin & Kilbinger 2015a.

Once ϵ_t is set, we draw a particle from the previous weighted set of particles $\vec{\theta}_{t-1}$. This particle is perturbed by a kernel, set to the covariance of $\vec{\theta}_{t-1}$. Then once again, we generate a simulation by forward modeling $X = f(\vec{\theta^i})$, calculate the distance $\rho(X, D)$, and compare the distance to the new distance threshold (ϵ_t) in order to accept or reject the particle. This process is repeated until we assemble a new set of N particles $\vec{\theta}_t$. We then update the particle weights according to the kernel, the prior distribution, and the previous set of weights, as described in Algorithm 1. The entire procedure is then repeated for the next iteration, $t+1$.

There are a number of ways to specify the perturbation kernel in the ABC-PMC algorithm. A widely used technique is to define the perturbation kernel as a multivariate Gaussian centered on the weighted mean of the particle population with a covariance matrix set to the covariance of the particle population. This perturbation kernel is often called the global multivariate Gaussian kernel. For a thorough discussion of various schemes for specifying the perturbation kernel, we refer the reader to Filippi et al. 2011.

The iterations continue in the ABC-PMC algorithm until convergence is confirmed. One way to ensure convergence is to impose a threshold for the acceptance ratio, which is measured in each iteration. The acceptance ratio is the ratio of the number of proposals accepted by the distance threshold, to the full number of proposed particles at every step. Once the acceptance ratio for an iteration falls below the imposed threshold, the algorithm has converged and is suspended. Another way to ensure convergence is by monitoring the fractional change in the distance threshold ($\epsilon_t/\epsilon_{t-1} - 1$) after each iteration. When the fractional change becomes smaller than some specified tolerance level, the algorithm has reached convergence. Another convergence criteria, is through the derived uncertainties of the inferred parameters measured after each iteration. When the uncertainties stabilize and show negligible variations, convergence is ensured. In Section 2.4.2 we detail the specific convergence criteria used in our analysis.

2.3.2 Forward model

2.3.2.1 Halo Occupation Modeling

ABC requires a forward generative model. In large scale structure studies, this implies a model that is able to generate a galaxy catalog. We then calculate and compare summary statistics of the data and model catalog in an identical fashion In this section, we describe the forward generative model we use within the framework of the halo occupation distribution.

The assumption that galaxies reside in dark matter halos is the bedrock underlying all contemporary theoretical predictions for galaxy clustering. The Halo Occupation Distribution (HOD) is one of the most widely used approaches to characterizing this galaxy-halo connection. The central quantity in the HOD is $p(N_{\rm g}|M_{\rm h})$, the probability that a halo of mass $M_{\rm h}$ hosts $N_{\rm g}$ galaxies.

The most common technical methods for estimating the theoretical galaxy 2PCF utilize the first two moments of P, which contain the necessary information to calculate the one- and two-halo terms of the galaxy correlation function:

$$1+\xi^{1h}_{\rm gg}(r) \simeq \frac{1}{4\pi r^2 \bar{n}^2_{\rm g}} \int {\rm d}M_{\rm h} \frac{{\rm d}n}{{\rm d}M_{\rm h}} \Xi_{\rm gg}(r|M_{\rm h}) \times \langle N_{\rm g}(N_{\rm g}-1)|M_{\rm h}\rangle \,, \tag{2.8}$$

and

$$\xi^{2h}_{\rm gg}(r) \simeq \xi_{\rm mm}(r) \left[\frac{1}{\bar{n}_{\rm g}} \int {\rm d}M_{\rm h} \frac{{\rm d}n}{{\rm d}M_{\rm h}} \langle N_{\rm g}|M_{\rm h}\rangle \, b_{\rm h}(M_{\rm h})\right]^2 \tag{2.9}$$

In Eqs. (2.8) and (2.9), $\bar{n}_{\mathrm{g}}$ is the galaxy number density, $\mathrm{d}n/\mathrm{d}M_{\mathrm{h}}$ is the halo mass function, the spatial bias of dark matter halos is $b_{\mathrm{h}}(M_{\mathrm{h}})$, and ξ_{mm} is the correlation function of dark matter. If we represent the spherically symmetric intra-halo distribution of galaxies by a unit-normalized $n_{\mathrm{g}}(r)$, then the quantity $\Xi_{\mathrm{gg}}(r)$ appearing in the above two equations is the convolution of $n_{\mathrm{g}}(r)$ with itself. These fitting functions are calibrated using N-body simulations.

Fitting function techniques, however, require many simplifying assumptions. For example, Eqs. (2.8) and (2.9) assume that the galaxy distribution within a halo is spherically symmetric. These equations also face well-known difficulties of properly treating halo exclusion and scale-dependent bias, which results in additional inaccuracies commonly exceeding the 10% level (van den Bosch et al. 2013). Direct emulation methods have made significant improvements in precision and accuracy in recent years (Heitmann et al. 2009, 2010); however, a labor- and computation-intensive interpolation exercise must be carried out each time any alternative statistic is explored, which is one of the goals of the present work.

To address these problems, throughout this paper we make no appeal to fitting functions or emulators. Instead, we use the `Halotools` package to populate dark matter halos with mock galaxies and then calculate our summary statistics directly on the resulting galaxy catalog with the same estimators that are used on observational data (Hearin et al. 2016a). Additionally, through our forward modeling approaching, we are able to explore observables beyond the 2PCF, such as the group multiplicity function, for which there is no available fitting function. This framework allows us to use group multiplicity function for providing quantitative constraints on the galaxy-halo connection. In the following section, we will show that using this observable, we can obtain constraints on the HOD parameters comparable to those found from the 2PCF measurements.

For the fiducial HOD used throughout this paper, we use the model described in Zheng et al. 2007a. The occupation statistics of central galaxies follow a nearest-integer distribution with first moment given by

$$\langle N_{\mathrm{cen}} \rangle = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log M - \log M_{\mathrm{min}}}{\sigma_{\log M}}\right)\right]. \tag{2.10}$$

Satellite occupation is governed by a Poisson distribution with the mean given by

$$\langle N_{\mathrm{sat}} \rangle = \langle N_{\mathrm{cen}} \rangle \left(\frac{M - M_0}{M_1}\right)^{\alpha}. \tag{2.11}$$

We assume that central galaxies are seated at the exact center of the host dark matter halo and are at rest with respect to the halo velocity, defined according to `Rockstar` halo finder (Behroozi et al. (2013d)) as the mean velocity of the inner 10% of particles in the halo. Satellite galaxies are confined to reside within the virial radius following an NFW spatial profile (Navarro et al. 2004) with a concentration parameter given by the $c(M)$ relation (Dutton & Macciò 2014). The peculiar velocity of satellites with respect to their host halo is calculated according to the solution of the Jeans equation of an NFW profile (More et al. 2009a). We refer the reader to Hearin et al. (2016b), Hearin et al. (2016a), and `http://halotools.readthedocs.io` for further details.

For the halo catalog of our forward model, we use the publicly available `Rockstar` (Behroozi et al. 2013d) halo catalogs of the `MultiDark` cosmological N-body simulation (Riebe et al. 2011).[2] `MultiDark` is a collision-less dark-matter only N-body simulation. The ΛCDM cosmological parameters of `MultiDark` are $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.042$, $n_s = 0.95$, $\sigma_8 = 0.82$, and $h = 0.7$. The gravity solver used in the N-body simulation is the Adaptive Refinement Tree code (ART; Kravtsov et al. 1997) run on 2048^3 particles in

[2] In particular, we use the `halotools_alpha_version2` version of this catalog, made publicly available as part of `Halotools`.

a 1 h^{-1}Gpc periodic box. `MultiDark` particles have a mass of $m_p \simeq 8.72 \times 10^8\ h^{-1}M_\odot$; the force resolution of the simulation is $\epsilon \simeq 7h^{-1}$ kpc.

One key detail of our forward generative model is that when we populate the `MultiDark` halos with galaxies, we do not populate the entire simulation volume. Rather, we divide the volume into a grid of 125 cubic subvolumes, each with side lengths of 200 h^{-1}Mpc. We refer to these subvolumes as $\{\texttt{BOX1}, ..., \texttt{BOX125}\}$. The first subvolume is reserved to generate the mock observations which we describe in Section 2.4.1. When we simulate a galaxy catalog for a given $\vec{\theta}$ in parameter space, we randomly select one of the subvolumes from $\{\texttt{BOX2}, ..., \texttt{BOX125}\}$ and then populate the halos within this subvolume with galaxies. We implement this procedure to account for sample variance within our forward generative model.

2.3.3 Summary Statistics

One of the key ingredients for parameter inference using ABC, is the distance metric between the data and the simulations. In essence, it quantifies how close the simulation is to reproducing the data. The data and simulation in our scenario (the HOD framework) are galaxy populations and their positions. A direct comparison, which would involve comparing the actual galaxy positions of the populations, proves to be difficult. Instead, a set of statistical summaries are used to encapsulate the information of the data and simulations. These quantities should sufficiently describe the information of the data and simulations while providing the convenience for comparison. For the positions of galaxies, sensible summary statistics, which we later use in our analysis, include

- Galaxy number density, $\bar{n}_\mathrm{g}$: the comoving number density of galaxies computed by dividing the comoving volume of the sample from the total number of galaxies. $\bar{n}_\mathrm{g}$ is measured in units of $(\mathrm{Mpc}/h)^{-3}$.

- Galaxy two-point correlation function, $\xi_{\rm gg}(r)$: a measurement of the excess probability of finding a galaxy pair with separation r over an random distribution. To compute $\xi_{\rm gg}(rr)$ in our analysis, for computational reasons, we use the Natural estimator (Peebles 1980a):
$$\xi(r) = \frac{DD}{RR} - 1, \tag{2.12}$$
where DD and RR refer to counts of data-data and random-random pairs.

- Galaxy group multiplicity function, $\zeta_{\rm g}(N)$: the number density of galaxy groups in bins of group richness N where group richness is the number of galaxies within a galaxy group. We rely on a Friends-of-Friends (hereafter FoF) group-finder algorithm (Davis et al. 1985a) to identify galaxy groups in our galaxy samples. That is, if the separation of a galaxy pair is smaller than a specified linking length, the two galaxies are assigned to the same group. The FoF group-finder has been used to identify and analyze the galaxy groups in the SDSS main galaxy sample (Berlind et al. (2006b)). For details regarding the group finding algorithm, we refer readers to Davis et al. (1985a).

  In this study we set the linking length to be 0.25 times the mean separation of galaxies which is given by $\bar{n}_{\rm g}^{-1/3}$. Once the galaxy groups are identified, we bin them into bins of group richness. The total number of groups in each bin is divided by the comoving volume to get $\zeta_{\rm g}(N)$ — in units of $({\rm Mpc}/h)^{-3}$.

2.4 ABC at work

With the methodology and the key components of ABC explained above, here we set out to demonstrate how ABC can be used to constrain HOD parameters. We start, in Section 2.4.1 by creating our "observation". We select a set of HOD parameters which we

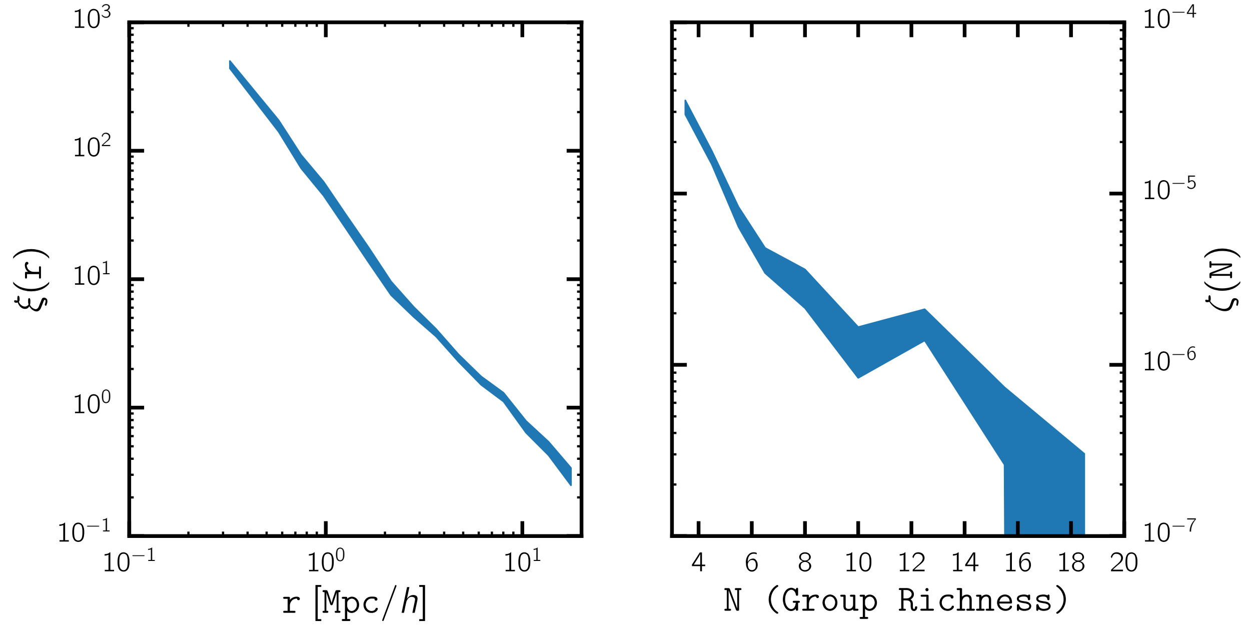

Figure 2.1: The two-point correlation function $\xi_{\rm gg}(r)$ (left) and group multiplicity function $\zeta_{\rm g}(N)$ (right) summary statistics of the mock observations generated from the "true" HOD parameters described in Section 2.4.1. The width of the shaded region corresponds to the square root of the covariance matrix diagonal elements (Eq. 2.14). In our ABC analysis, we treat the $\xi_{\rm gg}(r)$ and $\zeta_{\rm g}(N)$ above as the summary statistics of the observation.

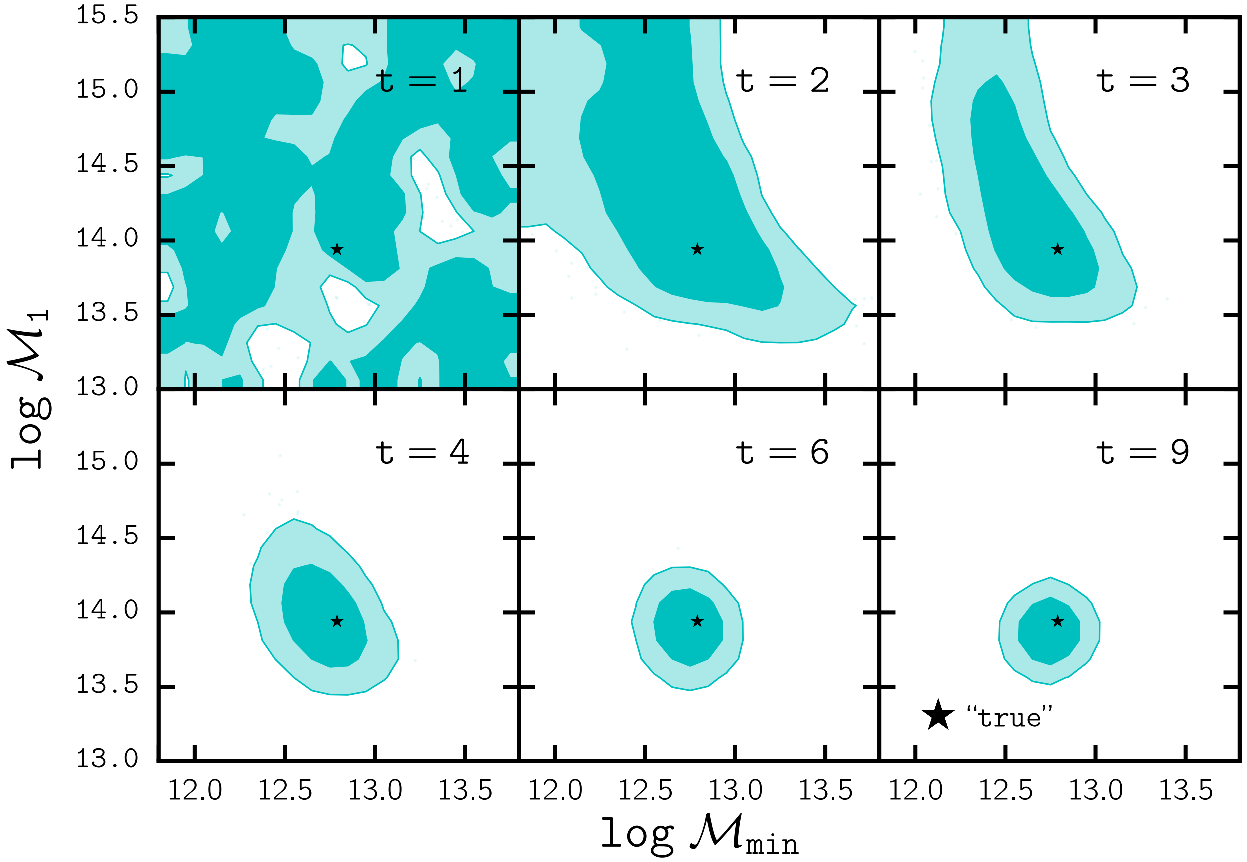

Figure 2.2: We demonstrate the evolution of the ABC particles, $\vec{\theta}_t$, over iterations $t = 1$ to 9 in the $\log \mathcal{M}_{min}$ and $\log \mathcal{M}_1$ parameter space. $\bar{n}$ and $\zeta_{\rm g}(N)$ are used as observables for the above results. For reference, in each panel, we include the "true" HOD parameters (black star) listed in Section 2.4.1. The initial distance threshold, $\vec{\epsilon}_1 = [\infty, \infty]$ at $t = 1$ (top left) so the $\vec{\theta}_1$ spans the entire range of the prior distribution, which is also the range of the panels. We see for $t < 5$, the parameter space occupied by the ABC $\vec{\theta}_t$ shrinks dramatically. Eventually when the algorithm converges, $t > 7$, the parameter space occupied by $\vec{\theta}_t$ no longer shrinks and their distributions represent the posterior distribution of the parameters. At $t = 9$, the final iteration, the ABC algorithm has converged and we find that $\vec{\theta}_{\rm true}$ lies safely within the 68% confidence region.

deem as the "true" parameters and run it through our forward model producing a catalog of galaxy positions which we treat as our observation. Then, in Section 2.4.2, we explain the distance metric and other specific choices we make for the ABC-PMC algorithm. Ultimately, we demonstrate the use of ABC in LSS, in Section 2.4.3, where we present the parameter constraints we get from our ABC analyses. Lastly, in order to both assess the quality of the ABC-PMC parameter inference and also discuss the assumptions of the standard Gaussian likelihood approach, we compare the ABC-PMC results to parameter constraints using the standard approach in Section 2.4.4.

2.4.1 Mock Observations

In generating our "observations", and more generally for our forward model, we adopt the HOD model from Zheng et al. (2007a) where the expected number of galaxies populating a dark matter halo is governed by Eqs (2.10) and (2.11). For the parameters of the model used to generate the fiducial mock observations, we choose the Zheng et al. (2007a) best-fit HOD parameters for the SDSS main galaxy sample with a luminosity threshold $M_r = -21$:

$\log M_{\rm min}$	$\sigma_{\log M}$	$\log M_0$	$\log M_1$	α
12.79	0.39	11.92	13.94	1.15

Since these parameters are used to generate the mock observation, they are the parameters that we ultimately want to recover from our parameter inference. We refer to them as the true HOD parameters. Plugging them into our forward model (Section 2.3.2), we generate a catalog of galaxy positions.

For our summary statistics of the catalogs we use: the mean number density $\bar{n}_{\rm g}$, the galaxy two-point correlation function $\xi_{\rm gg}(r)$, and the group multiplicity function $\zeta_{\rm g}(N)$. Our mock observation catalog has $\bar{n}_{\rm g} = 9.28875 \times 10^{-4}\ h^{-3}{\rm Mpc}^3$ and in Figure 2.1 we plot $\xi_{\rm gg}(r)$

(left panel) and $\zeta_{\rm g}(N)$ (right panel). The width of the shaded region represent the square root of the diagonal elements of the summary statistic covariance matrix, which is computed as we describe below.

We calculate $\xi_{\rm gg}$ using the natural estimator (Section 2.3.3) with fifteen radial bins. The edges of the first radial bin are 0.15 and 0.5 h^{-1}Mpc. The bin edges for the next 14 bins are logarithmically-spaced between 0.5 and 20 h^{-1}Mpc. We compute the $\zeta_{\rm g}(N)$ as described in Section 2.3.3 with nine richness bins where the bin edges are logarithmically-spaced between 3 and 20. To calculate the covariance matrix, we first run the forward model using the true HOD parameters for all 125 halo catalog subvolumes: $\{\texttt{BOX1}, ..., \texttt{BOX125}\}$. We compute the summary statistics of each subvolume galaxy sample k:

$$\mathbf{x}^{(k)} = [\bar{n}_{\rm g},\ \xi_{\rm gg},\ \zeta_{\rm g}], \tag{2.13}$$

Then we compute the covariance matrix as

$$\begin{aligned} {\rm C}^{\rm sample}_{i,j} &= \frac{1}{N_{\rm mocks}-1} \sum_{k=1}^{N_{\rm mocks}} \Big[\mathbf{x}^{(k)}_i - \overline{\mathbf{x}}_i\Big]\Big[\mathbf{x}^{(k)}_j - \overline{\mathbf{x}}_j\Big], & (2.14)\\ \text{where } \overline{\mathbf{x}}_i &= \frac{1}{N_{\rm mocks}} \sum_{k=1}^{N_{\rm mocks}} \mathbf{x}^{(k)}_i. & (2.15) \end{aligned}$$

Throughout our ABC-PMC analysis, we treat the $\bar{n}_{\rm g}$, $\xi_{\rm gg}(r)$, and $\zeta_{\rm g}(N)$ we describe in this section as if they were the summary statistics of actual observations. However, we benefit from the fact that these observables are generated from mock observations using the true HOD parameters of our choice: we can use the true HOD parameters to assess the quality of the parameter constraints we obtain from ABC-PMC.

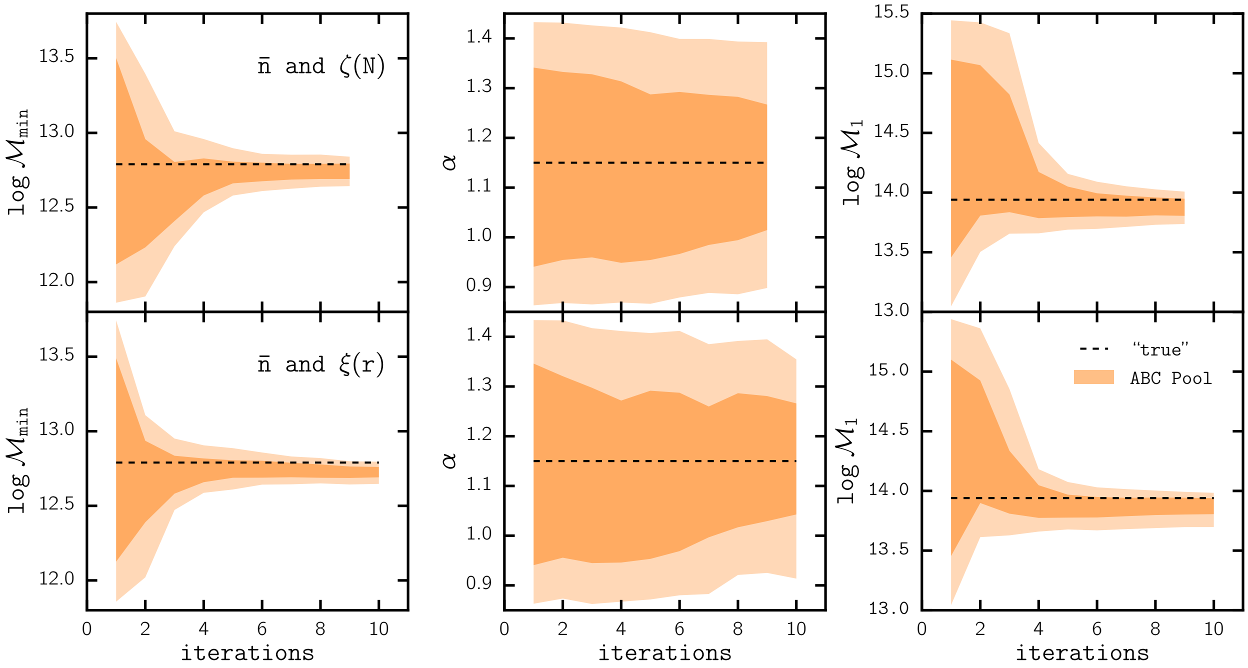

Figure 2.3: We illustrate the convergence of the ABC algorithm through the evolution of the ABC particle distribution as a function of iteration for parameters $\log \mathcal{M}_{\rm min}$ (left), α (center), and $\log \mathcal{M}_1$ (right). The top panel corresponds our ABC results using the observables $(\bar{n}, \zeta_{\rm g}(N))$, while the lower panel plots corresponds to the ABC results using $(\bar{n}, \xi_{\rm gg}(r))$. The distributions of parameters show no significant change after $t > 7$, which suggests that the ABC algorithm has converged.

2.4.2 ABC-PMC Design

In Section 4.6.1, we describe the key components of the ABC algorithm we use in our analysis. Now, we describe the more specific choices we make within the algorithm: the distance metric, the choice of priors, the distance threshold, and the convergence criteria. So far we have described three summary statistics: $\bar{n}_{\rm g}$, $\xi_{\rm gg}(r)$, and $\zeta_{\rm g}(N)$. In order to explore the detailed differences in the ABC-PMC parameter constraints based on our choice of summary statistics, we run our analysis for two sets of observables: ($\bar{n}_{\rm g}$, $\xi_{\rm gg}$) and ($\bar{n}_{\rm g}$, $\zeta_{\rm g}$).

For both analyses, we use a multi-component distance (Silk et al. 2012, Cisewsky et al in preparation). Each summary statistic has a distance associated to it: ρ_n, ρ_ξ, and ρ_ζ. We calculate each of these distance components as,

$$\begin{aligned}
\rho_n &= \frac{\left(\bar{n}_{\rm g}^{\rm d} - \bar{n}_{\rm g}^{\rm m}\right)^2}{\sigma_n^2}, && (2.16)\\
\rho_\xi &= \sum_k \frac{\left[\xi_{\rm gg}^{\rm d}(r_k) - \xi_{\rm gg}^{\rm m}(r_k)\right]^2}{\sigma_{\xi,k}^2}, && (2.17)\\
\rho_\zeta &= \sum_k \frac{\left[\zeta_{\rm g}^{\rm d}(N_k) - \zeta_{\rm g}^{\rm m}(N_k)\right]^2}{\sigma_{\zeta,k}^2}. && (2.18)
\end{aligned}$$

The superscripts d and m denote the data and model respectively. The data, are the observables calculated from the mock observation (Section 2.4.1). σ_n^2, $\sigma_{\xi,k}^2$, and $\sigma_{\zeta,k}^2$ are not the diagonal elements of the covariance matrix (2.14). Instead, they are diagonal elements of the covariance matrix $C^{\rm ABC}$.

We construct $C^{\rm ABC}$ by populating the entire `MultiDark` halo catalogs 125 times repeatedly, calculating $\bar{n}_{\rm g}$, $\xi_{\rm gg}$, and $\zeta_{\rm g}$ for each realization, and then computing the covariance associated with these observables across all realizations. We highlight that $C^{\rm ABC}$ differs from Eq. 2.14, in that it does not populate the 125 subvolumes but the entire `MultiDark` simulation and therefore does not incorporate sample variance. The ABC-PMC analysis instead accounts for the sample variance through the forward generative model, which populates the subvolumes in the same manner as the observations. We use σ_n^2, $\sigma_{\xi,k}^2$, and $\sigma_{\zeta,k}^2$ to ensure that the distance is not biased to variations of observables on specific radial or richness bin.

For our ABC-PMC analysis using the observables $\bar{n}_\mathrm{g}$ and ξ_gg, our distance metric $\vec{\rho} = [\rho_n, \rho_\xi]$ while the distance metric for the ABC-PMC analysis using the observables $\bar{n}_\mathrm{g}$ and ζ_g, is $\vec{\rho} = [\rho_n, \rho_\zeta]$. To avoid any complications from the choice for our prior, we select uniform priors over all parameters aside from the scatter parameter $\sigma_{\log\ M}$, for which we choose a log-uniform prior. We list the range of our prior distributions in Table 2.1.

With the distances and priors specified, we now describe the distance thresholds and the convergence criteria we impose in our analyses. For the initial iteration, we set distance thresholds for each distance component to ∞. This means, that the initial pool $\vec{\theta}_1$ is simply sampled from the prior distribution we specify above. After the initial iteration, the distance threshold is adaptively lowered in subsequent iterations. More specifically, we follow the choice of Lin & Kilbinger (2015a) and set the distance threshold $\vec{\epsilon}_t$ to the median of $\vec{\rho}_{t-1}$, the multi-component distance of the previous iteration of particles ($\vec{\theta}_{t-1}$).

The distance threshold $\vec{\epsilon}_t$ will progressively decrease. Eventually after a sufficient number of iterations, the region of parameter space occupied by $\vec{\theta}_t$ will remain unchanged. As this happens, the acceptance ratio begins to fall significantly. When the acceptance ratio drops below 0.001, our acceptance ratio threshold of choice, we deem the ABC-PMC algorithm as converged. In addition to the acceptance ratio threshold we impose, we also ensure that distribution of the parameters converges – another sign that the algorithm has converged. Next, we present the results of our ABC-PMC analyses using the sets of observables ($\bar{n}_\mathrm{g}$, ξ_gg) and ($\bar{n}_\mathrm{g}$, ζ_g).

Table 2.1: **Prior Specifications**: The prior probability distribution and its range for each of the Zheng et al. (2007a) HOD parameters. All mass parameters are in unit of $h^{-1}M_{\odot}$

HOD Parameter	Prior	Range
α	Uniform	[0.8, 1.3]
$\sigma_{\log \mathrm{M}}$	Log-Uniform	[0.1, 0.7]
$\log M_0$	Uniform	[10.0, 13.0]
$\log M_{min}$	Uniform	[11.02, 13.02]
$\log M_1$	Uniform	[13.0, 14.0]

2.4.3 Results: ABC

We describe the ABC algorithm in Section 4.6.1 and list the particular choices we make in the implementation in the previous section. Finally, we demonstrate how the ABC algorithm produces parameter constraints and present the results of our ABC analysis – the parameter constraints for the Zheng et al. (2007a) HOD model.

We begin with a qualitative demonstration of the ABC algorithm in Figure 2.2, where we plot the evolution of the ABC $\vec{\theta}_t$ over the iterations $t = 1$ to 9, in the parameter space of $[\log \mathcal{M}_1, \log \mathcal{M}_{min}]$. The ABC procedure we plot in Figure 2.2 uses $\bar{n}$ and $\zeta_{\mathrm{g}}(N)$ for observables, but the overall evolution is the same when we use $\bar{n}$ and $\xi_{\mathrm{gg}}(r)$. The darker and lighter contours represent the 68% and 95% confident regions of the posterior distribution over $\vec{\theta}_t$. For reference, we also plot the "true" HOD parameter $\vec{\theta}_{\mathrm{true}}$ (black star) in each of the panels. The parameter ranges of the panels are equivalent to the ranges of the prior probabilities we specify in Table 2.1.

For $t = 1$, the initial pool (top left), the distance threshold $\vec{\epsilon}_1 = [\infty, \infty]$, so $\vec{\theta}_1$ uniformly samples the prior probability over the parameters. At each subsequent iteration, the threshold is lowered (Section 2.4), so for $t < 6$ panels, we note that the parameter spaced occupied by $\vec{\theta}_t$ dramatically shrinks. Eventually when the algorithm begins to converge, $t > 7$, the contours enclosing the 68% and 95% confidence interval stabilize. At the final iteration $t = 9$ (bottom right), the algorithm has converged and we find that $\vec{\theta}_{\mathrm{true}}$ lies within the 68% confidence interval of the $\vec{\theta}_{t=9}$ particle distribution. This $\vec{\theta}_t$ distribution at the final iteration represents the posterior distribution of the parameters.

To better illustrate the criteria for convergence, in Figure 2.3, we plot the evolution of the $\vec{\theta}_t$ distribution as a function of iteration for parameters $\log \mathcal{M}_{\mathrm{min}}$ (left), α (center), and $\log \mathcal{M}_1$ (right). The darker and lighter shaded regions correspond to the 68% and 95% confidence levels of the $\vec{\theta}_t$ distributions. The top panels correspond to our ABC results using $(\bar{n}, \zeta_{\mathrm{g}})$ as observables and the bottom panels correspond to our results using $(\bar{n}, \xi_{\mathrm{gg}})$. For each of the parameters in both top and bottom panels, we find that the distribution does not evolve significantly for $t > 7$. At this point additional iterations in our ABC algorithm will neither impact the distance threshold $\vec{\epsilon}_t$ nor the posterior distribution of $\vec{\theta}_t$. We also emphasize that the convergence of the parameter distributions coincides with when the acceptance ratio, discussed in Section 2.4.2, crosses the predetermined shut-off value of 0.001. Based on these criteria, our ABC results for both $(\bar{n}, \zeta_{\mathrm{g}})$ and $(\bar{n}, \xi_{\mathrm{gg}})$ observables have converged.

We present the parameter constraints from the converged ABC analysis in Figure 2.4 and Figure 2.5. Figure 2.4 shows the parameter constraints using $\bar{n}$ and $\xi_{\mathrm{gg}}(r)$ while Figure 2.5 plots the constraints using $\bar{n}$ and $\zeta_{\mathrm{g}}(N)$. For both figures, the diagonal panels plot the posterior distribution of the HOD parameters with vertical dashed lines marking the 50% (median) and 68% confidence intervals. The off-diagonal panels plot the degeneracy between parameter pairs. To determine the accuracy of our ABC parameter constraints, we plot the "true" HOD parameters (black) in each of the panels. For both sets of observables, our ABC constraints are consistent with the "true" HOD parameters. For $\log \mathcal{M}_0$, $\log \sigma_{\log M}$, and α, the true parameter values lie near the center of the 68% confidence interval. For the other parameter, which have much tighter constraints, the true parameters lie within the 68% confidence interval.

To further test the ABC results, in Figure 2.6, we compare $\xi_{\rm gg}(r)$ (left) and $\zeta_{\rm g}(N)$ (right) of the mock observations from Section 2.4.1 to the predictions of the ABC posterior distribution (shaded). The error bars of the mock observations represent the square root of the diagonal elements of the covariance matrix (Eq. 2.14) while the darker and lighter shaded regions represent the 68% and 95% confidence regions of the ABC posterior predictions. In the lower panels, we plot the ratio of the ABC posterior prediction $\xi_{\rm gg}(r)$ and $\zeta_{\rm g}(N)$ over the mock observation $\xi_{\rm gg}^{\rm obvs}(r)$ and $\zeta_{\rm g}^{\rm obvs}(N)$. Overall, the ratio of the 68% confidence region of ABC posterior predictions is consistent with unity throughout the r and N range. We observe slight deviations in the $\xi_{\rm gg}$ ratio for $r > 5$ Mpc/h; however, any deviation is within the uncertainties of the mock observations. Therefore, the observables drawn from the ABC posterior distributions are in good agreement with the observables of the mock observation.

The ABC results we obtain using the algorithm of Section 4.6.1 with the choices of Section 2.4.2 produce parameter constraints that are consistent with the "true" HOD parameters (Figures 2.4 and 2.5). They also produce observables $\xi_{\rm gg}(r)$ and $\zeta_{\rm g}(N)$ that are consistent with $\xi_{\rm gg}^{\rm obvs}$ and $\zeta_{\rm g}^{\rm obvs}$. Thus, through ABC we are able to produce consistent parameter constraints. *More importantly, we demonstrate that ABC is feasible for parameter inference in large scale structure.*

2.4.4 Comparison to the Gaussian Pseudo - Likelihood MCMC Analysis

In order to assess the quality of the parameter inference described in the previous section, we compare the ABC-PMC results with the HOD parameter constraints from assuming a Gaussian likelihood function. The model used for the Gaussian likelihood analysis is different than the forward generative model adopted for the ABC-PMC algorithm, to be consistent with the standard approach.

In the ABC analysis, the model accounts for sample variance by randomly sampling a subvolume to be populated with galaxies. Instead, in the Gaussian pseudo-likelihood analysis, the covariance matrix is assumed to capture the uncertainties from sample variance. Hence, in the model for the Gaussian pseudo-likelihood analysis, we populate halos of the *entire* `MultiDark` simulation rather than a subvolume. We describe the Gaussian pseudo-likelihood analysis below.

To write down the Gaussian pseudo-likelihood, we first introduce the vector $\mathbf{x}$: a combination of the summary statistics (observables) for a galaxy catalog. When we use $\bar{n}_{\rm g}$ and $\xi_{\rm gg}(r)$ as observables in the analysis: $\mathbf{x} = [\bar{n}_{\rm g}, \xi_{\rm gg}]$; when we use $\bar{n}_{\rm g}$ and $\zeta_{\rm g}(N)$ as observables in the analysis: $\mathbf{x} = [\bar{n}_{\rm g}, \zeta_{\rm g}]$. Based on this notation, we can write pseudo-likelihood function as

$$-2 \ln \mathcal{L}(\theta|d) \quad = \quad \Delta\mathbf{x}^T \widehat{C^{-1}} \Delta\mathbf{x} + \ln\Big[(2\pi)^d \det(C)\Big], \tag{2.19}$$

where

$$\Delta\mathbf{x} \quad = \quad [\mathbf{x}_{obs} - \mathbf{x}_{mod}], \tag{2.20}$$

the difference between $\mathbf{x}_{obs}$, measured from the mock observation, and $\mathbf{x}_{mod}(\theta)$ measured from the mock catalog generated from the model with parameters θ . d here is the dimension of $\mathbf{x}$ (for $\mathbf{x} = [\bar{n}_{\rm g}, \xi_{\rm gg}]$, $d = 13$; for $\mathbf{x} = [\bar{n}_{\rm g}, \zeta_{\rm g}]$, $d = 10$). $\widehat{C^{-1}}$ is the inverse covariance matrix, which we estimate following Hartlap et al. (2007):

$$\widehat{C^{-1}} = \frac{N_{\rm mocks} - d - 1}{N_{\rm mocks} - 1}\, \widehat{C}^{-1}. \tag{2.21}$$

$\widehat{C}$ is the estimated covariance matrix, calculated using the corresponding $\mathbf{x}$ block of the covariance matrix from Eq. 2.14, and $N_{\rm mock}$ is the number of mocks used for the estimation ($N_{\rm mock} = 124$; see Section 2.4.1). We note that in $\widehat{C}$ the dependence on the HOD parameters is neglected, so the second term in the expression of Eq. 2.19 can be neglected. Finally, using this pseudo-likelihood, we sample from the posterior distribution given the prior distribution using the MCMC sampler `emcee` (Foreman-Mackey et al. 2013).

In Figures 2.7 and 2.8, we compare the results from ABC-PMC and Gaussian pseudo-likelihood MCMC analyses using $[\bar{n}_{\rm g}, \xi_{\rm gg}]$ and $[\bar{n}_{\rm g}, \zeta_{\rm g}]$ as observables, respectively. The top panels in each figure compares the marginalized posterior PDFs for three parameters of the HOD model: $\{\log \mathcal{M}_{\rm min}, \alpha, \log \mathcal{M}_1\}$. The lower panels in each figure compares the 68% and 95% confidence intervals of the constraints derived from the two inference methods as a box plot. The "true" HOD parameters are marked by vertical dashed lines in each panel.

In both Figures 2.7 and 2.8, the marginalized posteriors for each of the parameters from both inference methods are comparable and consistent with the "true" HOD parameters. However, we note that there are minor discrepancies between the maringalized posterior distributions. In particular, the distribution for α derived from ABC-PMC is less biased than the α constraints from the Gaussian pseudo-likelihood approach.

In Figures 2.9 and 2.10, we plot the contours enclosing the 68% and 95% confidence regions of the posterior probabilities of the two methods using $[\bar{n}_{\rm g}, \xi_{\rm gg}]$ and $[\bar{n}_{\rm g}, \zeta_{\rm g}]$ as observables respectively. In both figures, we mark the "true" HOD parameters (black star). The overall shape of the contours are in agreement with each other. However, we note that the contours for the ABC-PMC method are more extended along α.

Overall, the HOD parameter constraints from ABC-PMC are consistent with those from the Gaussian pseudo-likelihood MCMC method; however, using ABC-PMC has a number of advantages. For instance, ABC-PMC utilizes a forward generative model. Our forward generative model accounts for sample variance. On the other hand, the Gaussian pseudo-likelihood approach, as mentioned earlier this section, does not account for sample variance in the model and relies on the covariance matrix estimate to capture the sample variance of the data.

Accurate estimation of the covariance matrix in LSS, however, faces a number of challenges. It is both labor and computationally expensive and dependent on the accuracy of simulated mock catalogs, known to be unreliable on small scales (see Heitmann et al. 2008; Chuang et al. 2015a and references therein). In fact, as Sellentin & Heavens (2016) points out, using estimates of the covariance matrix in the Gaussian psuedo-likelihood approach become further problematic. Even when inferring parameters from a Gaussian-distributed data set, using covariance matrix estimates rather than the *true* covariance matrix leads to a likelihood function that is *no longer* Gaussian. ABC-PMC does not depend on a covariance matrix estimate; hence, it does not face these problems.

In addition to not requiring accurate covariance matrix estimates, forward models of the ABC-PMC method, in principle, also have the advantage that they can account for sources of systematic uncertainties that affect observations. All observations suffer from significant systematic effects which are often difficult to correct. For instance, in SDSS-III BOSS (Dawson et al., 2013b), fiber collisions and redshift failures siginifcantly bias measurements and analysis of observables such as $\xi_{\rm gg}$ or the galaxy powerspectrum (Ross et al., 2012; Guo et al., 2012b; Hahn et al., 2017). In parameter inference, these systematics can affect the likelihood, and thus any analysis that requires writing down the likelihood, in unknown ways. With a forward generative model of the ABC-PMC method, the systematics can be simulated and marginalized out to achieve unbiased constraints.

Furthermore, *ABC-PMC – unlike the Gaussian pseudo-likelihood approach – is agnostic about the functional form of the underlying distribution of the summary statistics* (*e.g.* $\xi_{\rm gg}$ and $\zeta_{\rm g}$). As we explain throughout the paper, the likelihood function in LSS *cannot* be

Gaussian. For $\xi_{\rm gg}$, the correlation function must satisfy non-trivial positive-definiteness requirements and hence the Gaussian pseudo-likelihood function assumption is not correct in detail. In the case of $\zeta_{\rm g}(N)$, assuming a Gaussian functional form for the likelihood, which in reality is more likely Poisson, misrepresents the true likelihood function. In fact, this incorrect likelihood, may explain why the constraints on α are less biased for the ABC-PMC analysis than the Gaussian-likelihood analysis in 2.10.

Although in our comparison using simple mock observations, we find generally consistent parameter constraints from both the ABC-PMC analysis and the standard Gaussian pseudo-likelihood analysis, more realistic scenarios present many factors that can generate inconsistencies. Consider a typical galaxy catalog from LSS observations. These catalogs consist of objects with different data qualities, signal-to-noise ratios, and systematic effects. For example, catalogs are often incomplete beyond some luminosity/redshift or have some threshold signal-to-noise ratio cut imposed on them.

These selection effects, coupled with the systematic effects earlier this section, make correctly predicting the likelihood intractable. In the standard Gaussian pseudo-likelihood analysis, and other analysis that require writing down a likelihood function, these effects can significantly bias the inferred parameter constraints. In these situations, employing ABC equipped with a generative forward model that incorproates selection and systematic effects may produce less biased parameter constraints.

Despite the advantages of ABC, one obstactle for adopting it to parameter inference has been the computational costs of generative forward models, a key element of ABC. By combining ABC with the PMC sampling method, however, ABC-PMC efficiently converges to give reliable posterior parameter constraints. In fact, in our analysis, the total computational resources required for the ABC-PMC analysis were *comparable* to the computational resources used for the Gaussian pseudo-likelihood analysis with MCMC sampling.

Applying ABC-PMC beyond the analysis in this work, to broader LSS analyses imposes some caveats. In this work, we focus on the galaxy-halo connection, so our generative forward model populates halos with galaxies. LSS analyses for inferring cosmological parameters would require generating halos by running cosmological simulations. The forward models also need to accurately model the observation systematic effects of the latest observations. Hence, accurate generative forward models in LSS analyses demand improvements in simulations and significant computational resources in order to infer unbiased parameter constraints. Recent cosmology simulations show promising improvements in both accuracy and speed (*e.g.* Feng et al., 2016). Such developements will be crucial for applying ABC-PMC to broader LSS analyses and exploiting the significant advantages that ABC-PMC offers.

2.5 Summary and Conclusion

Approximate Bayesian Computation, ABC, is a generative, simulation-based inference that can deliver correct parameter estimation with appropriate choices for its design. It has the advantage over the standard approach in that it does not require explicit knowledge of the likelihood function. It only relies on the ability to simulate the observed data, accounting for the uncertainties associated with observation and on specifying a metric for the distance between the observed data and simulation. When the specification of the likelihood function proves to be challenging or when the true underlying distribution of the observable is unknown, ABC provides a promising alternative for inference.

The standard approach to large scale structure studies relies on the assumption that the likelihood function for the observables – often two-point correlation function – given the model has a Gaussian functional form. In other words, it assumes that the statistical summaries are Gaussian distributed. In principle to rigorously test such an assumption, a large number of realistic simulations would need to be generated in order to examine the actual distribution of the observables. This process, however, is prohibitively—both labor and computationally —expensive. Therefore, our assumption of a Gaussian likelihood function remains largely unconfirmed and so unknown. Fortunately, the framework of ABC permits us to bypass any assumptions regarding the distribution of observables. Through ABC, we can provide constraints for our models without making the unexamined assumption of Gaussianity.

With the ultimate goal of demonstrating that ABC is feasible for LSS studies, we use it to constrain parameters of the halo occupation distribution, which dictates the galaxy-halo connection. We begin by constructing a mock observation of galaxy distribution with a chosen set of "true" HOD model parameters. Then we attempt to constrain these parameters using ABC. More specifically, in this paper:

- We provide an explanation of the ABC algorithm and present how Population Monte Carlo can be utilized to efficiently reach convergence and estimate the posterior distributions of model parameters. We use this ABC-PMC algorithm with a generative forward model built with `Halotools`, a software package for creating catalogs of galaxy positions based on models of the galaxy-halo connection such as the HOD.

- We choose $\bar{n}_{\rm g}$, $\xi_{\rm gg}$ and $\zeta_{\rm g}$ as observables and summary statistics of the galaxy position catalogs. And for our ABC-PMC algorithm, we specify a multi-component distance metric, uniform priors, a median threshold implementation, and an acceptance rate-based convergence criterion.

- From our specific ABC-PMC method, we obtain parameter constraints that are consistent with the "true" HOD parameters of our mock observations. Hence we demonstrate that ABC-PMC can be used for parameter inference in LSS studies.

- We compare our ABC-PMC parameter constraints to constraints using the standard Gaussian-likelihood MCMC analysis. The constraints we get from both methods are comparable in accuracy and precision. However, for our analysis using $\bar{n}_{\mathrm{g}}$ and ζ_{g} in particular, we obtain less biased posterior distributions when comparing to the "true" HOD parameters.

Based on our results, we conclude that ABC-PMC is able to consistently infer parameters in the context of LSS. We also find that the computation required for our ABC-PMC and standard Gaussian-likelihood analyses are comparable. Therefore, with the statistical advantages that ABC offers, we present ABC-PMC as an improved alternative for parameter inference.

Acknowledgements

We thank Jessie Cisewsky for reading and making valuable comments on the draft. We would also like to thank Michael R. Blanton, Jeremy R. Tinker, Uros Seljak, Layne Price, Boris Leidstadt, Alex Malz, Patrick McDonald, and Dan Foreman-Mackey for productive and insightful discussions. MV was supported by NSF grant AST-1517237. DWH was supported by NSF (grants IIS-1124794 and AST-1517237), NASA (grant NNX12AI50G), and the Moore-Sloan Data Science Environment at NYU. KW was supported by NSF grant AST-1211889. Computations were performed using computational resources at NYU-HPC. We thank Shenglong Wang, the administrator of NYU-HPC computational facility, for his consistent and continuous support throughout the development of this project. We would like to thank the organizers of the AstroHackWeek 2015 workshop (`http://astrohackweek.org/2015/`), since the direction and the scope of this investigation was —to some degree— initiated through discussions in this workshop. Throughout this in-

vestigation, we have made use of publicly available software packages `emcee` and `abcpmc`. We have also used the publicly available python implementation of the FoF algorithm `pyfof` (`https://github.com/simongibbons/pyfof`).

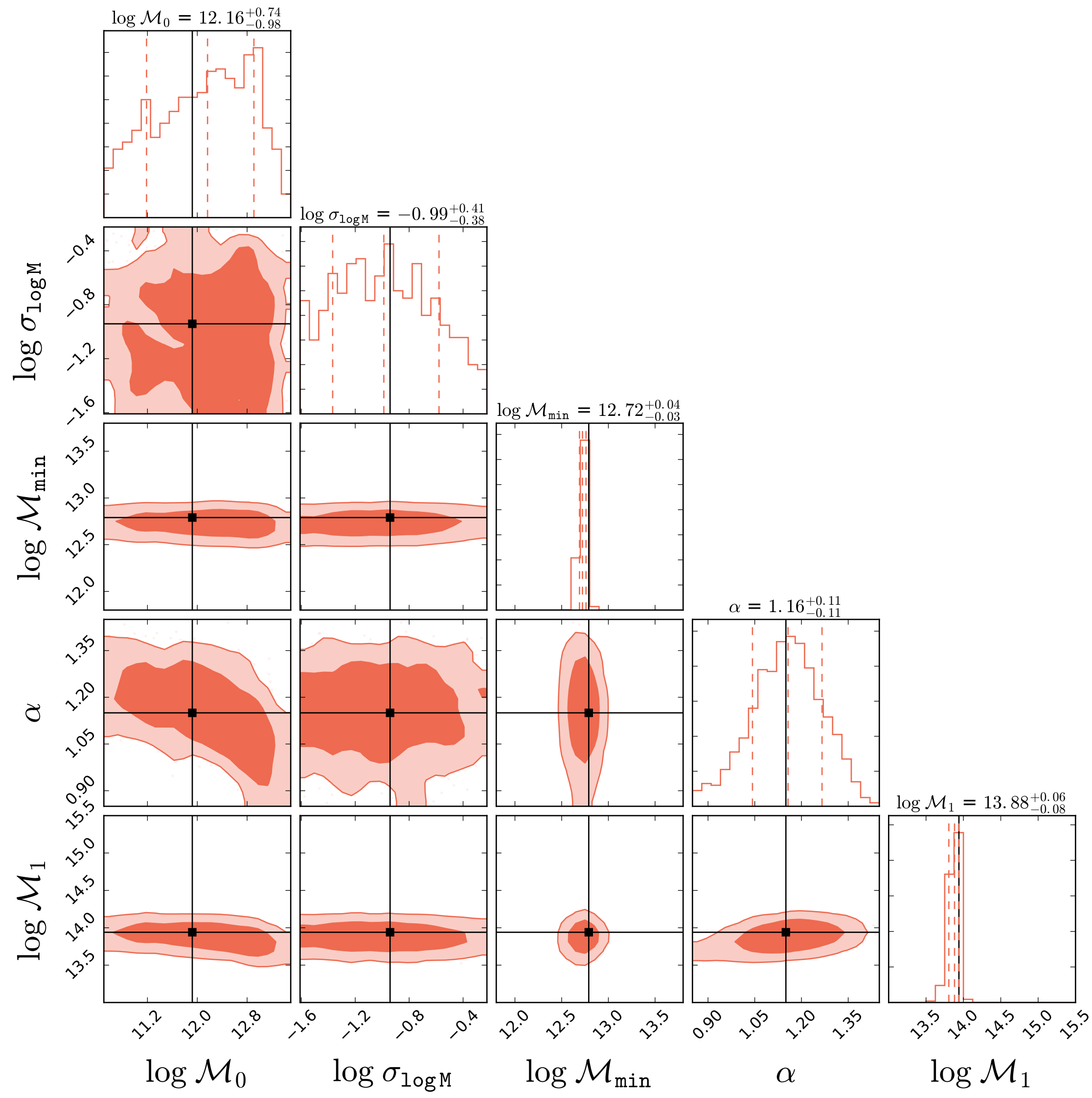

Figure 2.4: We present the constraints on the Zheng et al. (2007a) HOD model parameters obtained from our ABC-PMC analysis using $\bar{n}$ and $\xi_{\rm gg}(r)$ as observables. The diagonal panels plot the posterior distribution of each HOD parameter with vertical dashed lines marking the 50% quantile and 68% confidence intervals of the distribution. The off-diagonal panels plot the degeneracies between parameter pairs. The range of each panel corresponds to the range of our prior choice. The "true" HOD parameters, listed in Section 2.4.1, are also plotted in each of the panels (black). For $\log \mathcal{M}_0$, α, and $\sigma_{\log M}$, the "true" parameter values lie near the center of the 68% confidence interval of the posterior distribution. For $\log \mathcal{M}_1$ and $\log \mathcal{M}_{\rm min}$, which have tight constraints, the "true" values lie within the 68% confidence interval. Ultimately, the ABC parameter constraints we obtain in our analysis are consistent with the "true" HOD parameters.

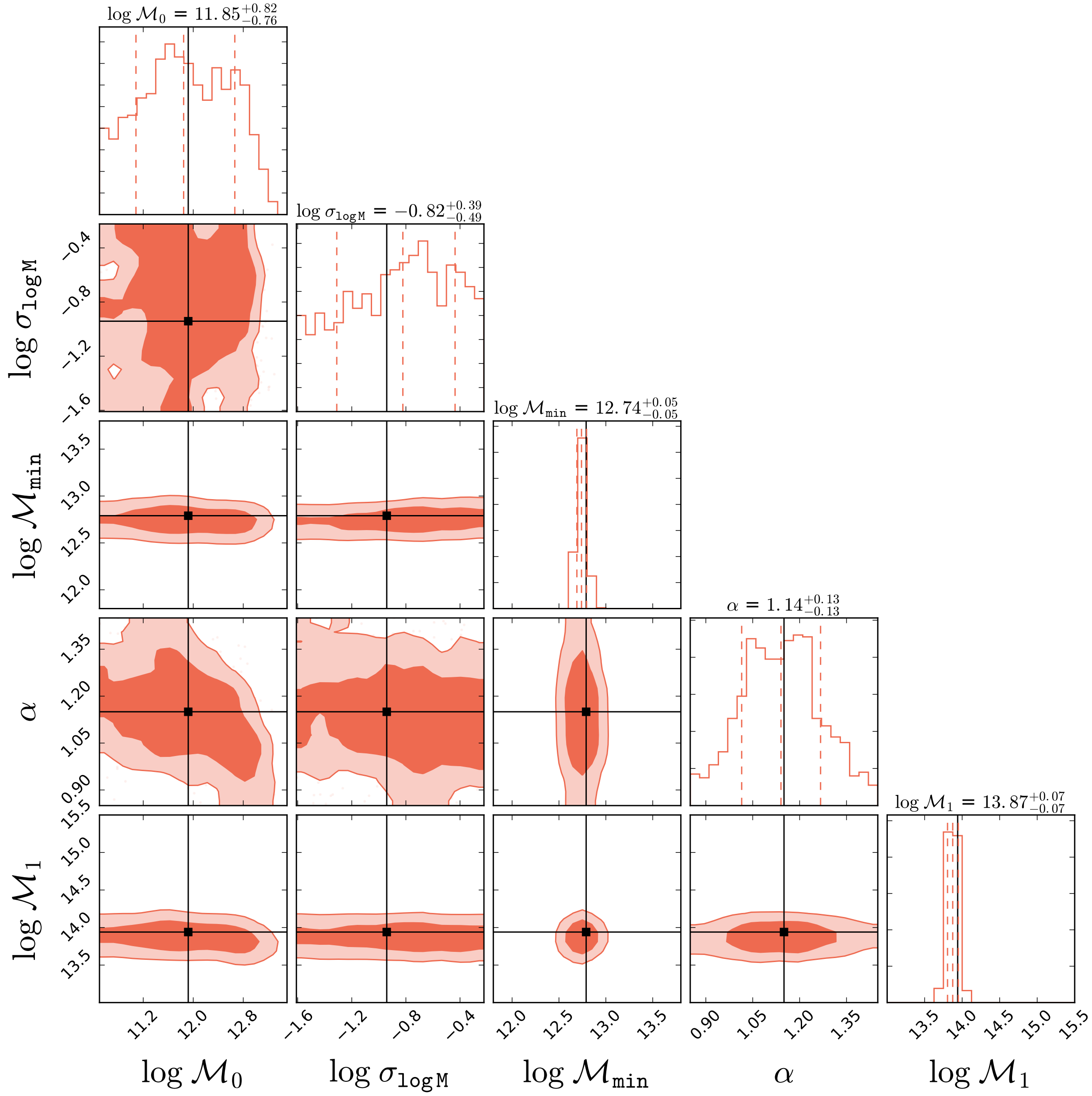

Figure 2.5: Same as Figure 2.4 but for our ABC analysis using $\bar{n}$ and $\zeta_{\rm g}(N)$ as observables. The ABC parameter constraints we obtain are consistent with the "true" HOD parameters.

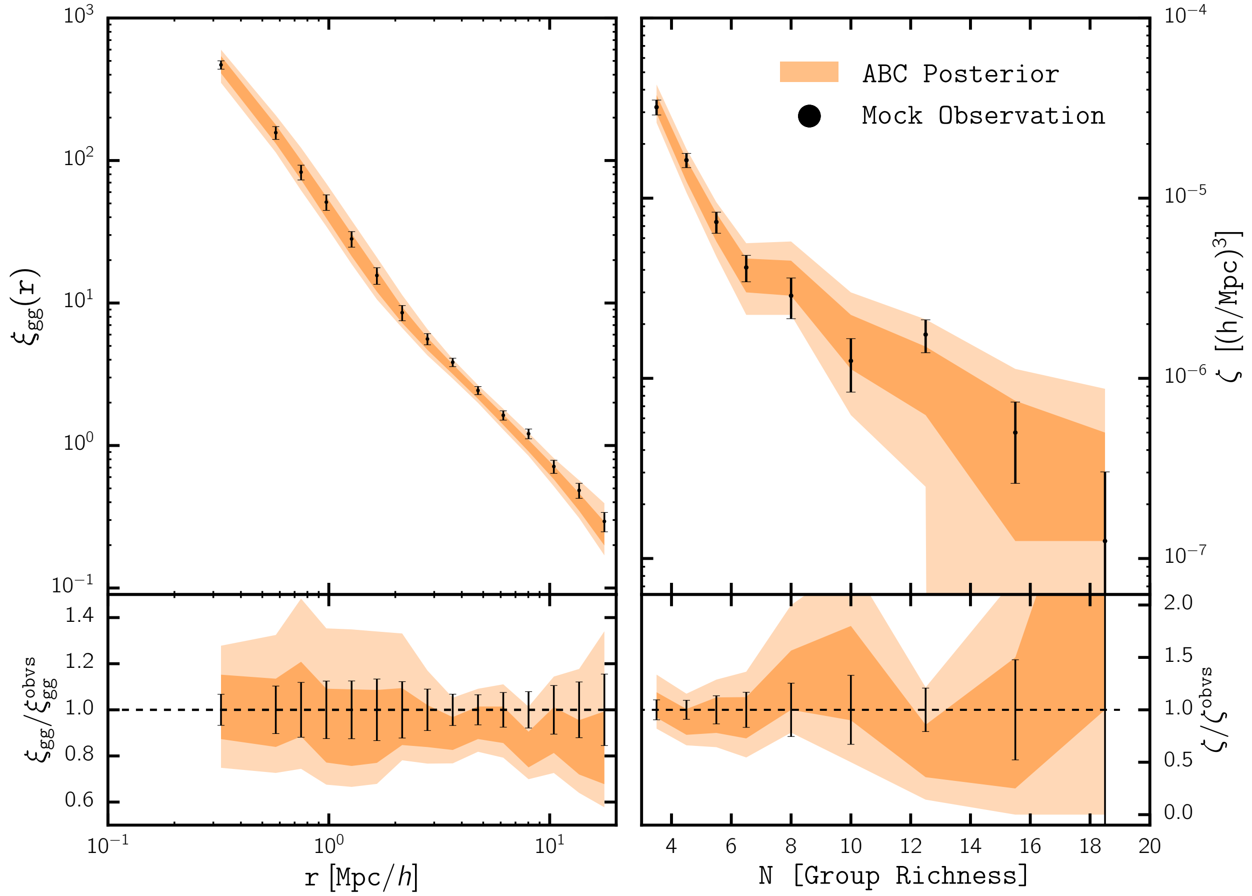

Figure 2.6: We compare the ABC-PMC posterior prediction for the observables $\xi_{gg}(r)$ (left) and $\zeta_g(N)$ (right) (orange; Section 2.4.3) to $\xi_{gg}(r)$ and $\zeta_g(N)$ of the mock observation (black) in the top panels. In the lower panels, we plot the ratio between the ABC-PMC posterior predictions for ξ_{gg} and ζ_g to the mock observation ξ_{gg}^{obvs} and ζ_g^{obvs}. The darker and lighter shaded regions represent the 68% and 95% confidence regions of the posterior predictions, respectively. The error-bars represent the square root of the diagonal elements of the error covariance matrix (equation 2.14) of the mock observations. Overall, the observables drawn from the ABC-PMC posteriors are in good agreement with ξ_{gg} and ζ_g of the mock observations. The lower panels demonstrate that for both observables, the error-bars of the mock observations lie within the 68% confidence interval of the ABC-PMC posterior predictions.

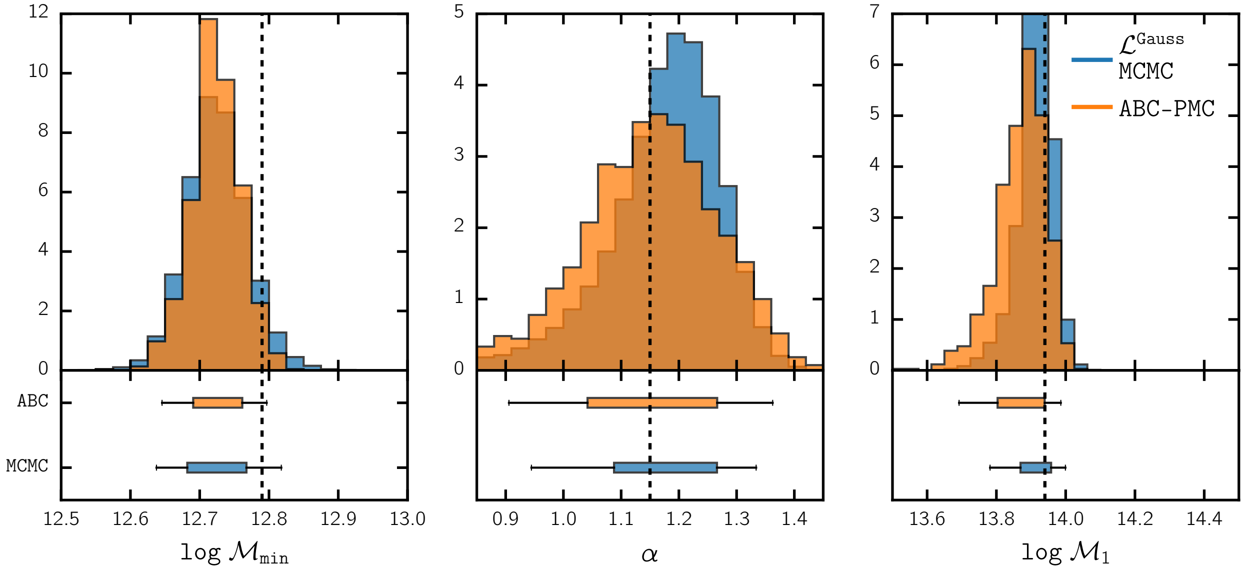

Figure 2.7: We compare the $\log \mathcal{M}_{\rm min}$, α, and $\log \mathcal{M}_1$ parameter constraints from ABC-PMC (orange) to constraints from the Gaussian pseudo-ikelihood MCMC (blue) using $\bar{n}_{\rm g}$ and $\xi_{\rm gg}(r)$ as observables. The *top* panels compares the two methods' marginalized posterior PDFs over the parameters. In the *bottom* panels, we include box plots marking the confidence intervals of the posterior distributions. The boxes represent the 68% confidence interval while the "whiskers" represent the 95% confidence interval. We mark the "true" HOD parameters with vertical black dashed line. The marginalized posterior PDFs obtained from the two methods are consistent with each other. The ABC-PMC and Gaussian pseudo-likelihood constraints are generally consistent for $\log \mathcal{M}_{\rm min}$ and $\log \mathcal{M}_1$. The ABC-PMC constraint for α is slightly less biased and has slightly larger uncertainty then the constraint from Gaussian pseudo-likelihood analysis.

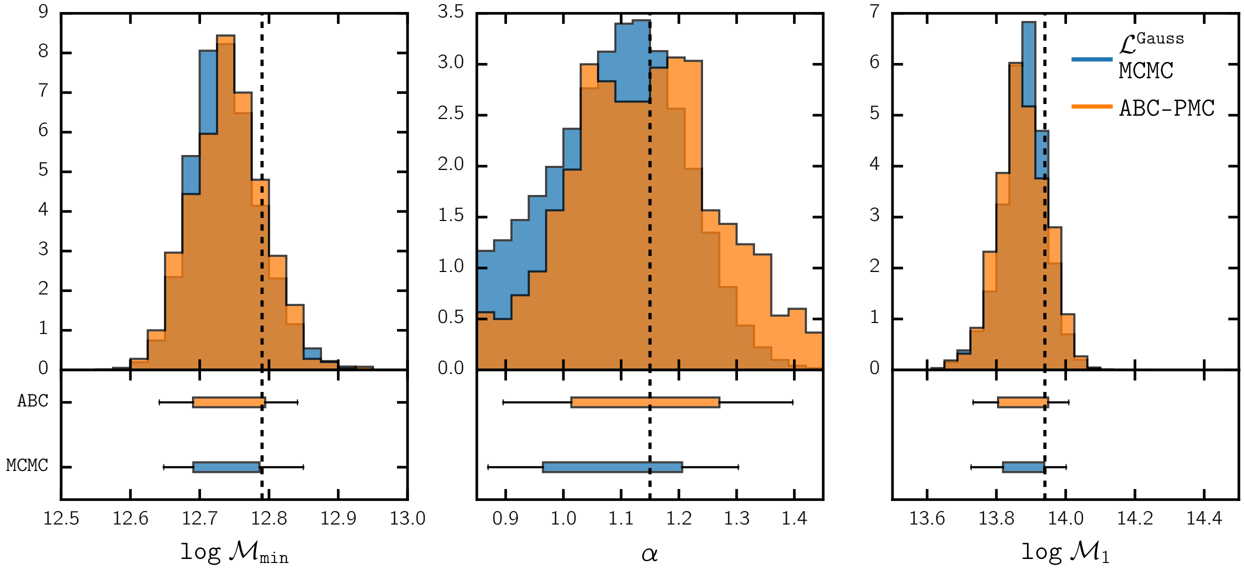

Figure 2.8: Same as Figure 2.7, but both the ABC-PMC analysis and the Gaussian pseudo-likelihood MCMC analysis use $\bar{n}_{\rm g}$ and $\zeta_{\rm g}(N)$ as observables. Both methods derive constraints consistent with the "true" HOD parameters and infer the region of allowed values to similar precision. We note that the MCMC constraint on α is slightly more biased compared to ABC-PMC estimate. This discrepancy may stem from the fact that the use of Gaussian pseudo-likelihood and its associated assumptions is more spurious when modeling the group multiplicity function.

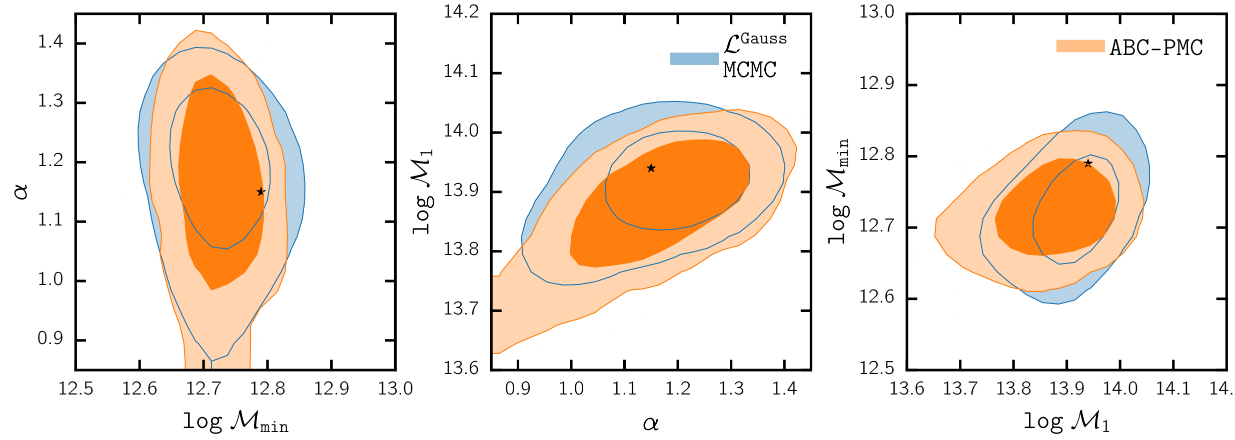

Figure 2.9: We compare the ABC-PMC (orange) and the Gaussian pseudo-likelihood MCMC (blue) predictions of the 68% and 95% posterior confidence regions over the HOD parameters ($\log \mathcal{M}_{\rm min}$, α, and $\log \mathcal{M}_1$) using $\bar{n}_{\rm g}$ and $\xi_{\rm gg}(r)$ as observables. In each panel, the black star represents the "true" HOD parameters used to generate the mock observations. Both inference methods derive confidence regions consistent with the "true" HOD parameters.

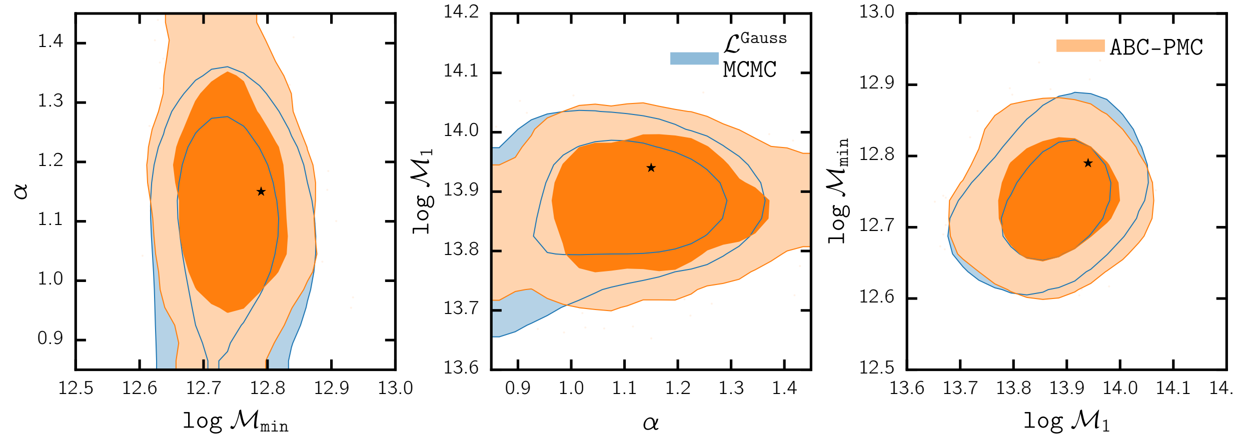

Figure 2.10: Same as Figure 2.9, but using $\bar{n}_{\rm g}$ and $\zeta_{\rm g}(N)$ as observables. Again, the confidence regions derived from both methods are consistent with the "true" HOD parameters used to generate the mock observations. The confidence region of α from the Gaussian pseudo-likelood method is biased compared to the ABC-PMC contours. This may be due to the fact that the true likelihood function that describes $\zeta_{\rm g}(N)$ deviates significantly from the assumed Gaussian functional form.

Chapter 3

PRIMUS: Effects of Galaxy Environment on the Quiescent Fraction Evolution at $z< 0.8$

This Chapter is joint work with Michael R. Blanton (NYU), John Moustakas (Siena College), Alison L. Coil (UCSD), Richard J. Cool (University of Arizona), Daniel J. Eisenstein (Harvard), Ramin A. Skibb (UCSD), Kenneth C. Wong (University of Arizona), and Guangtun Zhu (Johns Hopkins Unviersity) published in *The Astrophysical Jounral* as Hahn et al. (2015).

3.1 Chapter Abstract

We investigate the effects of galaxy environment on the evolution of the quiescent fraction (f_{Q}) from $z = 0.8$ to 0.0 using spectroscopic redshifts and multi-wavelength imaging data from the PRIsm MUlti-object Survey (PRIMUS) and the Sloan Digitial Sky Survey (SDSS).

Our stellar mass limited galaxy sample consists of $\sim 14,000$ PRIMUS galaxies within $z = 0.2 - 0.8$ and $\sim 64,000$ SDSS galaxies within $z = 0.05 - 0.12$. We classify the galaxies as quiescent or star-forming based on an evolving specific star formation cut, and as low or high density environments based on fixed cylindrical aperture environment measurements on a volume-limited environment defining population. For quiescent and star-forming galaxies in low or high density environments, we examine the evolution of their stellar mass function (SMF). Then using the SMFs we compute $f_{\mathrm{Q}}(\mathcal{M}_*)$ and quantify its evolution within our redshift range. We find that the quiescent fraction is higher at higher masses and in denser environments. The quiescent fraction rises with cosmic time for all masses and environments. At a fiducial mass of $10^{10.5} M_\odot$, from $z \sim 0.7$ to 0.1, the quiescent fraction rises by 15% at the lowest environments and by 25% at the highest environments we measure. These results suggest that for a minority of galaxies their cessation of star formation is due to external influences on them. In other words, in the recent Universe a substantial fraction of the galaxies that cease forming stars do so due to internal processes.

3.2 Introduction

Galaxies, in their detailed properties, carry the imprints of their surroundings, with a strong dependence of the quiescent fraction of galaxies on their local environment (e.g. Hubble 1936a; Oemler 1974; Dressler 1980; Hermit et al. 1996; Guzzo et al. 1997; for a recent review see Blanton & Moustakas 2009b). The strength of this dependence is itself a strongly decreasing function of galaxy stellar mass; at the extreme, the lowest mass ($< 10^9$ $M_\odot$) galaxies end their star formation only in dense regions, and never in isolation (Geha et al. 2012a). These effects also vary with redshift at least in the densest clusters, as observed in the changing fraction of late-type spirals relative to the field, found in studies of the morphology-density relation (Dressler 1984; Fasano et al. 2000; Smith et al. 2005; Desai et al. 2007). Clearly understanding the properties of galaxies in the present-day universe requires a careful investigation of the role of environment, and how that role changes over time.

Nevertheless, the evolution of the role of environment is a relatively subtle effect and must be interpreted within the context of the evolving galaxy population. For instance, the most dramatic change in galaxy properties during the past eight billion years has been the remarkable decline in the star-formation rate of galaxies in the Universe (Hopkins & Beacom 2006a). This decline appears dominated by decreases in the rates of star-formation of individual galaxies (Noeske et al. 2007). There is evidence that a large fraction of the decline is associated with strongly infrared-emitting starbursts (Bell et al. 2005; Magnelli et al. 2009). As Cooper et al. (2008) and others have pointed out, because the environmental dependence of total star-formation rates at fixed redshift is relatively small, environmental effects are unlikely to cause the overall star-formation rate decline.

During this period, the major classes of galaxies that we observe today have already been firmly in place (Bundy et al. 2006b; Borch et al. 2006a; Taylor et al. 2009; Moustakas et al. 2013). Though not as dramatic as the history of galaxies prior to $z \sim 1$, detailed observations of the stellar mass function find significant evolution of the galaxy population with the decline in the number density of massive star-forming galaxies accompanied by an increase in the number density of quiescent galaxies (Blanton 2006; Bundy et al. 2006b; Borch et al. 2006a; Moustakas et al. 2013). Moustakas et al. (2013), for instance, find that since $z \sim 1.0$ the $\sim 50\%$ decline in the number density of massive star-forming galaxies ($\mathcal{M}_* > 10^{11}\mathcal{M}_\odot$) is complemented by the rise in number density of intermediate-mass quiescent galaxies ($\mathcal{M}_* \approx 10^{9.5} - 10^{10}\mathcal{M}_\odot$), by a factor of $2 - 3$, and massive quiescent galaxies ($\mathcal{M}_* > 10^{11}\mathcal{M}_\odot$), by $\sim 20\%$. On the color-magnitude diagram, this corresponds to the doubling of the red sequence over this period (Bell et al. 2004; Borch et al. 2006a; Faber et al. 2007). These changes in galaxy population are likely a result of physical processes that cause the cessation of star-formation in star-forming galaxies.

Of the numerous mechanisms that have been proposed to explain this cessation, favored models suggest that internal processes such as supernovae or active galactic nuclei heat the gas within the galaxy, which consequently suppresses the cold gas supply used for star-formation (Kereš et al. 2005; Croton et al. 2006; Dekel & Birnboim 2008). Other models propose that environment dependent external processes such as ram-pressure stripping (Gunn & Gott 1972; Bekki 2009), strangulation (Larson et al. 1980; Balogh et al. 2000), or harassment (Moore et al. 1998) contribute to the cessation.

Observations such as Weinmann et al. (2006) and Peng et al. (2010) credit some of these proposed internal processes for the cessation of star-formation, especially in massive galaxies. Meanwhile, observations of galaxy properties such as color and morphology correlating with environment suggest that environment may play a role in ceasing star-formation (Blanton & Moustakas 2009b and references therein). However, it remains to be determined whether the environmental trends in galaxy properties reflect the direct effect of external environment on the galaxies' evolution (e.g. ram pressure, tidal forces, mergers) or reflect statistical differences in the histories of galaxies in different environments (e.g. an earlier formation time in dense regions).

In this paper we take the most straightforward investigation by directly determining the star-forming properties of galaxies as a function of environment, stellar mass and redshift in a single, consistently analyzed data set. This analysis can reveal how galaxies end their star formation over time, quantitatively establish the contribution of environmental effects to the overall trends, and reveal whether those trends happen equally in all environments. However, such an analysis has not been done previously due to the lack of sufficiently large samples.

In this paper, we apply this approach using the PRIism MUlti-object Survey (PRIMUS; Coil et al. 2011, Cool et al. 2013), the largest available redshift survey covering the epochs between $0 < z < 1$.

In Section 3.3 we present a brief description of the PRIMUS and SDSS data, our mass complete sample construction, and galaxy environment measurements. After dividing our galaxy sample into subsamples of star-forming or quiescent and high or low density environments, we compute and examine the evolution of the stellar mass functions for our subsamples in Section 3.4. In Section 3.5, we calculate the quiescent fraction, analyze the evolution of the quiescent fraction, quantify the effects of environment on the quiescent fraction evolution, and discuss the implications of our quiescent fraction results on the cessation of star-formation in galaxies. Finally in Section 4.8 we summarize our results.

Throughout the paper we assume a cosmology with $\Omega_m = 0.3, \Omega_\Lambda = 0.7$, and $H_0 = 70\,\mathrm{km\ s^{-1}Mpc^{-1}}$. All magnitudes are AB-relative.

3.3 Sample Selection

We are interested in quantifying the effects of galaxy environment on the evolution of the quiescent fraction over the redshift range $0 < z < 1$. For our analysis, we require a sample with sufficient depth and high quality spectroscopic redshift to probe the redshift range and to robustly measure galaxy environment. PRIMUS with its $\sim 120,000$ spectroscopic redshifts provides a large data set at intermediate redshifts for our analysis. In addition, we anchor our analysis with a low redshift sample derived from the Sloan Digital Sky Survey (York et al. 2000).

In Section 3.3.1 and Section 4.3 we provide a brief summary of the PRIMUS and SDSS data used for our sample selection. In Section 3.3.3 we define our stellar mass complete galaxy

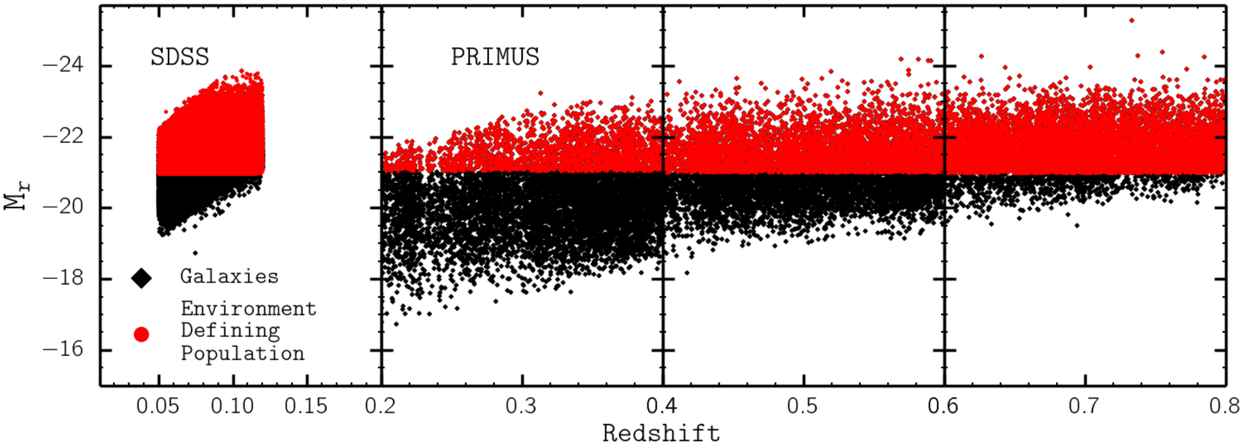

Figure 3.1: Absolute magnitude M_r versus redshift for our mass complete galaxy sample (black squares) with the Environment Defining Population (red circles) plotted on top. Both samples are divided into redshift bins: $z = 0.05 - 0.12$, $0.2 - 0.4$, $0.4 - 0.6$, and $0.6 - 0.8$ (panels left to right). The lowest redshift bin ($z \approx 0.05 - 0.12$; leftmost panel) contain our galaxy sample and EDP selected from SDSS. The rest contain galaxies and EDP selected from PRIMUS. The redshift limits for the lowest redshift bin are empirically selected based on the bright and faint limits of SDSS galaxies. Stellar mass completeness limits, described in Section 3.3.3, are imposed on the galaxy population. Meanwhile, M_r limits are applied to the EDP such that the number density in each panel are equivalent (Section 3.3.5).

sample. Then, in Section 3.3.4, we classify the sample galaxies as quiescent or star-forming. We calculate the environment using a volume-limited Environment Defining Population in Section 3.3.5. Finally, in Section 3.3.6, we account for edge effects in the surveys.

3.3.1 PRIMUS

At intermediate redshifts we use multiwavelength imaging and spectroscopic redshifts from PRIMUS, a faint galaxy survey with $\sim 120,000$ redshifts ($\sigma_z/(1+z) \approx 0.5\%$) within the range $z \approx 0 - 1.2$. The survey was conducted using the IMACS spectrograph on the Magellan I Baade 6.5-m telescope with a slitmask and low dispersion prism. For details on the PRIMUS observation methods such as survey design, targeting, and data summary, we refer readers to Coil et al. (2011). For details on redshift fitting, redshift precision and survey completeness we refer readers to Cool et al. (2013).

While the PRIMUS survey targeted seven distinct extragalactic deep fields for a total of ~ 9 deg^2, we restrict our sample to five fields that have *GALEX* and *Spitzer*/IRAC imaging for a total of ~ 5.5 deg^2 (similar to the sample selection in Moustakas et al. 2013). Four of these fields are a part of the *Spitzer* Wide-area Infrared Extragalactic Survey (SWIRE[1]): the European Large Area ISO Survey - South 1 field (ELAIS-S1[2]), the Chandra Deep Field South SWIRE field (CDFS), and the XMM Large Scale Structure Survey field (XMM-LSS). The XMM-LSS consists of two separate but spatially adjacent fields: the Subaru/XMM-Newton DEEP Survey field (XMM-SXDS[3]) and the Canadian-France-Hawaii Telescope Legacy Survey field (XMM-CFHTLS[4]). Our fifth and final field is the Cosmic Evolution Survey (COSMOS[5]) field. For all of our fields we have near-UV (NUV) and far-UV (FUV) photometry from the *GALEX* Deep Imaging Survey (DIS; Martin et al. 2005; Morrissey et al. 2005) as well as ground-based optical and *Spitzer*/IRAC mid-infrared photometric catalogs. Moustakas et al. (2013) provides detailed descriptions of integrated flux calculations in the photometric bands for each of our fields. Furthermore, we derive the *K-corrections* from the photometry using `K-correct` (v4.2; Blanton & Roweis 2007).

Finally, using the spectroscopic redshift and broad wavelength photometry we apply `iSEDfit`, a Bayesian SED modeling code, to calculate stellar masses and star formation rates (SFRs) for our sample galaxies (Moustakas et al. 2013). `iSEDfit` uses the redshift and the observed photometry of the galaxies to determine the statistical likelihood of a large

[1] http://swire.ipac.caltech.edu/swire/swire.html

[2] http://dipastro.pd.astro.it/esis

[3] http://www.naoj.org/cience/SubaruProject/SDS

[4] http://www.cfht.hawaii.edu/Science/CFHLS

[5] http://cosmos.astro.caltech.edu

ensemble of generated model SEDs. The model SEDs are generated using Flexible Stellar Population Synthesis (FSPS) models (Conroy & Gunn 2010) based on the Chabrier (2003) IMF, along with a time dependent dust attenuation curve of Charlot & Fall (2000) and other prior parameters discussed in Section 4.1 and Appendix A of Moustakas et al. (2013). For details on the effects of prior parameter choices of iSEDfit on physical properties of galaxies we refer readers to the Appendix of Moustakas et al. (2013). For the observed photometry, we use the *GALEX* FUV and NUV, the two shortest IRAC bands at 3.6 and 4.5μm (the two longer-wavelength IRAC channels are excluded because `iSEDfit` does not model hot dust or polycyclic aromatic hydrocarbons emission lines), and the optical bands.

3.3.2 SDSS-GALEX

At low redshifts, we use spectroscopic redshifts and *ugriz* photometry from the SDSS Data Release 7 (DR7; Abazajian et al. 2009). More specifically we select galaxies from the New York University Value-Added Galaxy Catalog (hereafter VAGC) that satisfy the main sample criterion and have galaxy extinction corrected Petrosian magnitudes $14.5 < r < 17.6$ and spectroscopic redshifts within $0.01 < z < 0.2$ (Blanton et al. 2005b). We further restrict the VAGC sample to only galaxies with medium depth observations with total exposure time greater than 1 ks from *GALEX* Release 6. This leaves $167,727$ galaxies.

Next, we use the MAST/CasJobs[6] interface and a $4''$ diameter search radius, to obtain the NUV and FUV photometry for the SDSS-*GALEX* galaxies. For optical photometry, we use the *ugriz* bands from the SDSS `model` magnitudes scaled to the r-band `cmodel` magnitude. These photometric bands are then supplemented with integrated JHK_s magnitudes from the 2MASS Extended Source Catalog (XSC; Jarrett et al. 2000) and with photometry at 3.4 and

[6] http://galex.stsci.edu/casjobs

4.6μm from the WISE All-Sky Data Release[7]. Further details regarding the SDSS-*GALEX* sample photometry can be found in Section 2.4 of Moustakas et al. (2013). As previously done on the PRIMUS data in Section 3.3.1, we use `iSEDfit` to obtain the stellar masses and star formation rates for the SDSS-*GALEX* sample.

The SDSS-*GALEX* data discussed above is derived from the NYU-VAGC based on SDSS Data Release 7, using the standard SDSS photometric measurements. Several investigators have found that the background subtraction techniques used in the standard photometric catalogs introduce a size dependent bias in the galaxy fluxes and consequently stellar masses (West 2005; Blanton et al. 2005a; Lauer et al. 2007; Bernardi et al. 2007; Hyde & Bernardi 2009; West et al. 2010).

In order to quantify the effects of these photometric underestimations in our analysis, we tried replacing our SDSS fluxes in the *ugriz* band with *ugriz* fluxes from the NASA-Sloan Atlas (NSA) catalog, which incorporate the improved background subtraction presented in Blanton et al. (2011) and uses single-Seric fit fluxes rather than the standard SDSS `cmodel` fluxes. Using the ratio of the luminosity derived from the improved photometry over the luminosity derived from the standard NYU-VAGC photometry, we apply a preliminary correction to the stellar mass values obtained from `iSEDfit` assuming a consistent mass-to-light ratio. This mass correction leads to a significant increase in the stellar mass function for $\mathcal{M} > 10^{11}\mathcal{M}_{\odot}$; however, the effect of the mass correction was negligible for the quiescent fraction evolution results. As a result, for the results presented here we use the standard SDSS fluxes and we do not discuss the issues with photometric measurements any further in this paper. We note that a thorough investigation of these issues to understand their effect on the stellar mass function requires a reanalysis of both the SDSS photometry and the deeper photometry used for PRIMUS targeting.

[7] http://wise2.ipac.caltech.edu/docs/release/allsky

Table 3.1: Galaxy Subsamples

	$n_{\rm env}$	$N_{\rm gal}$		$\mathcal{M}_{\rm lim}$	
		Quiescent	Star-Forming	Quiescent	Star-Forming
$0.05 < z < 0.12$	$n_{\rm env} = 0.0$	6533	7508	$10^{10.2}\mathcal{M}_{\odot}$	$10^{10.2}\mathcal{M}_{\odot}$
	$n_{\rm env} > 3.0$	14673	9717		
$M_{{\rm r},lim} = -20.95$	all	33553	29864		
$0.2 < z < 0.4$	$n_{\rm env} = 0.0$	363	1231	$10^{9.8}\mathcal{M}_{\odot}$	$10^{9.8}\mathcal{M}_{\odot}$
	$n_{\rm env} > 3.0$	379	756		
$M_{{\rm r},lim} = -21.03$	all	1086	2879		
$0.4 < z < 0.6$	$n_{\rm env} = 0.0$	536	1498	$10^{10.3}\mathcal{M}_{\odot}$	$10^{10.3}\mathcal{M}_{\odot}$
	$n_{\rm env} > 3.0$	490	854		
$M_{{\rm r},lim} = -20.98$	all	1560	3577		
$0.6 < z < 0.8$	$n_{\rm env} = 0.0$	567	1254	$10^{10.7}\mathcal{M}_{\odot}$	$10^{10.6}\mathcal{M}_{\odot}$
	$n_{\rm env} > 3.0$	498	671		
$M_{{\rm r},lim} = -20.97$	all	1668	2964		
Total		77151			

Notes: Number of galaxies ($N_{\rm gal}$) in the mass complete subsamples within the edges of the survey (Section 3.3). The subsamples are classified based on environment ($n_{\rm env}$) and star formation rate (star-forming or quiescent). The lowest redshift bin is derived from SDSS; the rest are from PRIMUS. We also list the stellar mass completeness limit, $\mathcal{M}_{\rm lim}$, for our sample along with the r-band absolute magnitude limits, $M_{{\rm r},lim}$, for the Environment Defining Population.

3.3.3 Stellar Mass Complete Galaxy Sample

From the low redshift SDSS-*GALEX* and intermediate redshift PRIMUS data we define our mass complete galaxy sample. We begin by imposing the parent sample selection criteria from Moustakas et al. (2013). More specifically, we take the statistically complete *primary* sample from the PRIMUS data (Coil et al. 2011) and impose magnitude limits on optical selection bands as specified in Moustakas et al. (2013) Table 1. These limits are in different optical selection bands and have distinct values for the five PRIMUS target fields. We then exclude stars and broad-line AGN to only select objects spectroscopically classified as galaxies, with high-quality spectroscopic redshifts ($Q \geq 3$). Lastly, we impose a redshift range of $0.2 < z < 0.8$ for the PRIMUS galaxy sample, where $z > 0.2$ is selected due to limitations from sample variance and $z < 0.8$ is selected due to the lack of sufficient statistics in subsamples defined below.

For the PRIMUS objects that meet the above criteria, we assign statistical weights (described in Coil et al. 2011 and Cool et al. 2013) in order to correct for targeting incompleteness and redshift failures. The statistical weight, w_i, for each galaxy is given by

$$w_i = (f_{\text{target}} \times f_{\text{collision}} \times f_{\text{success}})^{-1}, \tag{3.1}$$

as in Equation (1) in Moustakas et al. (2013).

Since we are ultimately interested in a mass complete galaxy sample to derive SMFs and QFs, next we impose stellar mass completeness limits to our galaxy sample. Stellar mass completeness limits for a magnitude-limited survey such as PRIMUS are functions of redshift, the apparent magnitude limit of the survey, and the typical stellar mass-to-light ratio of galaxies near the flux limit. We use the same procedure as Moustakas et al. (2013), which follows Pozzetti et al. (2010), to empircally determine the stellar mass completeness limits.

For each of the target galaxies we compute $\mathcal{M}_{\rm lim}$ using $\log\ \mathcal{M}_{\rm lim} = \log\ \mathcal{M} + 0.4\ (m - m_{\rm lim})$, where $\mathcal{M}$ is the stellar mass of the galaxy in $\mathcal{M}_{\odot}$, $\mathcal{M}_{\rm lim}$ is the stellar mass of each galaxy if its magnitude was equal to the survey magnitude limit, m is the observed apparent magnitude in the selection band, and $m_{\rm lim}$ is the magnitude limit for our five fields. We construct a cumulative distribution of $\mathcal{M}_{\rm lim}$ for the 15% faintest galaxies in $\Delta z = 0.04$ bins. In each of these redshift bins, we calculate the minimum stellar mass that includes 95% of the galaxies. Separately for quiescent and star-forming galaxies, we fit quadratic polynomials to the minimum stellar masses versus redshift (star-forming or quiescent classification is described in the following section). Finally, we use the polynomials to obtain the minimum stellar masses at the center of redshift bins, $0.2-0.4$, $0.4-0.6$, and $0.6-0.8$, which are then used as PRIMUS stellar mass completeness limits.

For the low redshift portion of our galaxy sample, we start by limiting the SDSS-*GALEX* data to objects within $0.05 < z < 0.12$, a redshift range later imposed on the volume-limited Environment Defining Population (Section 3.3.5). To account for the targeting incompleteness of the SDSS-*GALEX* sample, we use the statistical weight estimates provided by the NYU-VAGC catalog. Furthermore, we determine a uniform stellar mass completeness limit of $10^{10.2}\mathcal{M}_{\odot}$ above the stellar mass-to-light ratio completeness limit of the SDSS-*GALEX* data within the imposed redshift limits (Blanton et al. 2005a; Baldry et al. 2008; Moustakas et al. 2013). We then apply this mass limit in order to obtain our mass-complete galaxy sample at low redshift.

We now have a stellar mass complete sample derived from SDSS-*GALEX* and PRIMUS data. Since our sample is derived from two different surveys, we account for the disparity in the redshift uncertainty. While PRIMUS provides a large number of redshifts out to $z = 1$, due to its use of a low dispersion prism, the redshift uncertainties are significantly larger ($\sigma_z/(1+z) \approx 0.5\%$) than the uncertainties of the SDSS redshifts. In order to have comparable environment measures throughout our redshift range, we apply PRIMUS redshift uncertainties to our galaxy sample selected from SDSS-*GALEX*. For each SDSS-*GALEX* galaxy, we adjust its redshift by randomly sampling a Gaussian distribution with standard deviation $\sigma = 0.005(1 + z_{\mathrm{SDSS}})$, where z_{SDSS} is the SDSS redshift of the galaxy.

3.3.4 Classifying Quiescent and Star-Forming Galaxies

We now classify our mass complete galaxy sample into quiescent or star-forming using an evolving cut based on specific star-formation rate utilized in Moustakas et al. (2013) Section 3.2. This classification method uses the star-forming (SF) sequence, which is the correlation between star-formation rate (SFR) and stellar mass in star-forming galaxies observed at least until $z \sim 2$ (Noeske et al. 2007; Williams et al. 2009; Karim et al. 2011). The PRIMUS sample displays a well-defined SF sequence within the redshift range of our galaxy sample. Using the power-law slope for the SF sequence from Salim et al. (2007) (SFR $\propto \mathcal{M}^{0.65}$) and the minimum of the quiescent/star-forming bimodality, determined empirically, we obtain the following equation to classify the target galaxies (Equation 2 in Moustakas et al. 2013):

$$\log(\mathrm{SFR}_{\mathrm{min}}) = -0.49 + 0.64\log(\mathcal{M} - 10) + 1.07(z - 0.1), \tag{3.2}$$

where $\mathcal{M}$ is the stellar mass of the galaxy. If the target galaxy SFR and stellar mass lie above Equation 3.2 we classify it as star-forming; if below, as quiescent (Moustakas et al. 2013 Figure 1.).

3.3.5 Galaxy Environment

We define the environment of a galaxy as the number of neighboring Environment Defining Population galaxies (defined below) within a fixed aperture centered around it. We

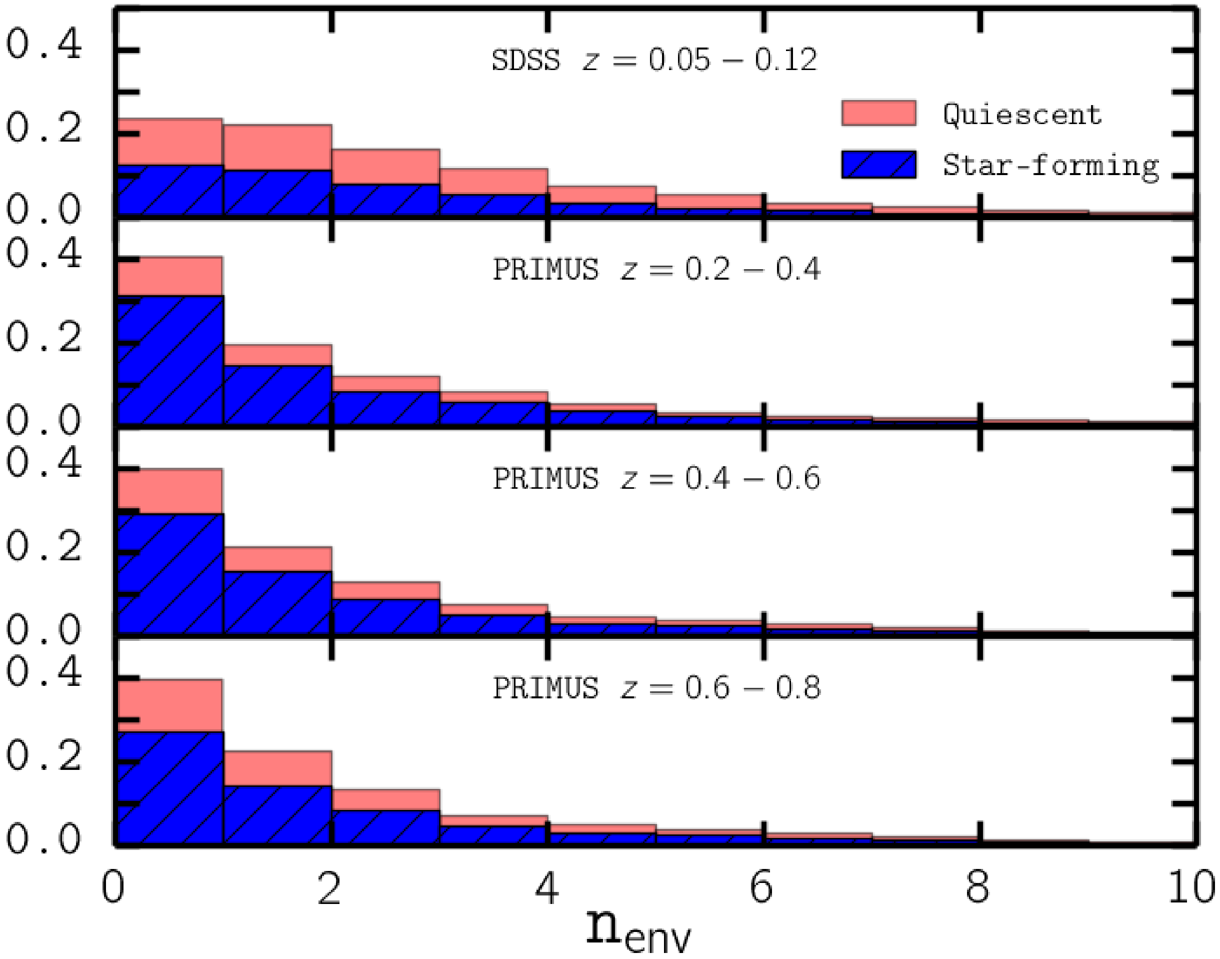

Figure 3.2: Normalized distribution of environment measurements ($n_{\rm env}$) for our mass complete galaxy sample within the survey edges. A fixed cylindrical aperture of $R_{\rm ap} = 2.5$ Mpc and $H_{\rm ap} = 35$ Mpc is used to measure environment. The star-forming galaxies contribution to the distribution is colored in blue and diagonally patterned. The contribution from quiescent galaxies is colored in red. Galaxies with $n_{\rm env} = 0.0$ are in low density environments and galaxies with $n_{\rm env} > 3.0$ are in high density environment. We note that the significant difference among the SDSS distribution and the PRIMUS distributions above is due to the different stellar mass completeness limits imposed on each redshift bin of our galaxy sample.

use fixed aperture measurements in order to quantify galaxy environment with an aperture sufficiently large to encompass massive halos (Muldrew et al. 2012; Skibba et al. 2013).

For our aperture, we use a cylinder of dimensions: $R_{\rm ap} = 2.5$ Mpc and $H_{\rm ap} = 35$ Mpc. We note that $H_{\rm ap}$ is the full height of the cylinder and $R_{\rm ap}$ and $H_{\rm ap}$ are comoving distances. We use a cylindrical aperture to account for the PRIMUS redshift errors and redshift space distortions (i.e. "Finger of God" effect). As Cooper et al. (2005) and Gallazzi et al. (2009) find, ± 1000 km $\rm s^{-1}$ optimally reduces the effects of redshift space distortions. The PRIMUS redshift uncertainty at $z \sim 0.7$ corresponds to $\sigma_z < 0.01$, so our choice of 35 Mpc for the aperture height accounts for both of these effects. Our choice of cylinder radius was motivated by scale dependence analyses in the literature (Blanton 2006; Wilman et al. 2010; Muldrew et al. 2012), which suggest that galactic properties such as color and quiescent fractions are most strongly dependent on scales < 2 Mpc, around the host dark matter halo sizes.

Before we measure the environment for our galaxy sample, we first construct a volume limited Environment Defining Population (EDP) with absolute magnitude cut-offs (M_r) in redshift bins with $\Delta z \sim 0.2$. The M_r cut-offs for the EDP are selected such that the cumulative number density over M_r for all redshift bins are equal. We make this choice in order to construct an EDP that contains similar galaxy populations through the redshift range (i.e. accounts for the progenitor bias). In their analysis of this method, Behroozi et al. (2013c) and Leja et al. (2013) find that although it does not precisely account for the scatter in mass accretion or galaxy-galaxy mergers, it provides a reasonable means to compare galaxy populations over a wide range of cosmic time.

In constructing the PRIMUS EDP we use the same PRIMUS data used to select our galaxy sample (described in Section 3.3.3). We again restrict the PRIMUS galaxies to $0.2 < z < 0.8$ and divide them into bins of $\Delta z = 0.2$. Before we consider the cumulative number densities in the redshift bins, we first determine the M_r limit for the highest redshift bin ($z = 0.6 - 0.8$) by examining the M_r distribution with bin size $\Delta M_r = 0.25$ and select $M_{r,\mathrm{lim}}$ near the peak of the distribution where bins with $M_r > M_{r,\mathrm{lim}}$ have fewer galaxies than the bin at $M_{r,\mathrm{lim}}$. We conservatively choose $M_{r,\mathrm{lim}}(0.6 < z < 0.8)$ to be $M_r = -20.97$. Then for the lower redshift bins, we impose absolute magnitude limits ($M_{r,\mathrm{lim}}$) such that the cumulative number density, calculated with the galaxy statistical weights, of the bin ordered by M_r is equal to the cumulative number density of the highest redshift bin with $M_{r,\mathrm{lim}}(0.6 < z < 0.8) = -20.97$.

For the SDSS EDP, we do not use the SDSS-*GALEX* parent data, which is limited to the combined angular selection window of the VAGC and *GALEX* (Section 4.3). Instead, since FUV, NUV values are not necessary for the EDP, we extend the parent data of the SDSS EDP to the entire NYU-VAGC, including galaxies outside of the *GALEX* window function. Furthermore, we impose a redshift range of $0.05 - 0.12$ on the SDSS EDP. This redshift range was determined to account for the lack of faint galaxies at $z \sim 0.2$ and the lack of bright galaxies at $z \sim 0.01$ in the VAGC. As with the PRIMUS redshift bins, we determine the SDSS EDP $M_{r,\mathrm{lim}}$ by matching the cumulative number density of the highest redshift bin. For redshift bins $z = 0.05 - 0.12$, $0.2 - 0.4$, $0.4 - 0.6$, $0.6 - 0.8$ we get $M_{r,\mathrm{lim}} = -20.95$, -21.03, -20.98 and -20.97, respectively. These absolute magnitude limits are illustrated in Figure 3.1, where we present the absolute magnitude (M_r) versus redshift for the galaxy sample (black squares) ad the EDP (red circles). The left-most panel corresponds to the samples derived from the SDSS-*GALEX* data while the rest correspond to samples derived from the PRIMUS data divided in bins with $\Delta z \sim 0.2$. Figure 3.1 shows clear M_r cutoffs in the M_r distribution versus redshift for the EDP on top of our galaxy sample.

For our SDSS-*GALEX* galaxy sample, in Section 3.3.3, we apply PRIMUS redshift errors in order to establish a consistent measurement of environment throughout our redshift range.

We appropriately apply equivalent redshift adjustments for the SDSS EDP. For the SDSS EDP galaxies that are also contained within the SDSS-*GALEX* sample, we adjust the redshift by an identical amount. For the rest, we apply the same redshift adjustment procedure described in Section 3.3.3 in order to obtain PRIMUS level redshift uncertainties.

Finally, we measure the environment for each galaxy in our galaxy sample by counting the number of EDP galaxies, $n_{\rm env}$, with RA, Dec, and z within our cylindrical aperture centered around it. $n_{\rm env}$ accounts for the statistical weights of the EDP galaxies. For our galaxy sample, the expected $n_{\rm env}$ given the uniform number density in each of our EDP redshift bin and volume of our cylindrical aperture is $\langle n_{\rm env} \rangle = 1.3$. Once we obtain environment measurements for all the galaxies in our galaxy sample, we classify galaxies with $n_{\rm env} = 0.0$ to be in "low" environment densities and galaxies with $n_{\rm env} > 3$ to be in "high" environment densities. The high environment cutoff was selected in order to reduce contamination from galaxies in low environment densities while maintaining sufficient statistics. In Section 3.5.2 we will also explore higher density cutoffs for $n_{\rm env}$.

The analysis we describe below uses a fixed cylindrical aperture with dimensions $R_{\rm ap} =$ 2.5 Mpc and $H_{\rm ap} = 35$ Mpc to measure environment. The same analysis was extended for varying aperture dimensions $R_{\rm ap} = 1.5,\ 2.5,\ 3.0$ Mpc and $H_{\rm ap} = 35,\ 70$ Mpc with adjusted environment classifications. The results obtained from using different apertures and environment classifications are qualitatively consistent with the results presented below.

3.3.6 Edge Effects

One of the challenges in obtaining accurate galaxy environments using a fixed aperture method is accounting for the edges of the survey. For galaxies located near the edge of the survey, part of the fixed aperture encompassing it will lie outside the survey regions. In this scenario, n_{env} only reflects the fraction of the environment within the survey geometry.

To account for these edge effects, we use a Monte Carlo method to impose edge cutoffs on our galaxy sample. First, using `ransack` from Swanson et al. (2008), we construct a random sample of $N_{\mathrm{ransack}} = 1,000,000$ points with RA and Dec randomly selected within the window function of the EDP (SDSS EDP and PRIMUS EDP separately). We then compute the angular separation, $\theta_{i,\mathrm{ap}}$ that corresponds to R_{ap} (Section 3.3.5) at the redshift of each sample galaxy i. For each sample galaxy we count the number of `ransack` points within $\theta_{i,\mathrm{ap}}$ of the galaxy: $n_{i,\mathrm{ransack}}$. Afterwards, we compare $n_{i,\mathrm{ransack}}$ to the expected value computed from the angular area of the environment defining aperture and the EDP window function:

$$\langle n_{\mathrm{ransack}} \rangle_i = \frac{N_{\mathrm{ransack}}}{A_{\mathrm{EDP}}} \times \pi \theta_{i,\mathrm{ap}}^2 \times f_{\mathrm{thresh}}. \tag{3.3}$$

A_{EDP} is the total angular area of the EDP window function and f_{thresh} is the fractional threshold for the edge effect cut-off. For $R_{\mathrm{ap}} = 2.5$ Mpc, we use $f_{\mathrm{thresh}} = 0.75$. If $n_{i,\mathrm{ransack}} > \langle n_{\mathrm{ransack}} \rangle_i$ then galaxy i remains in our sample; otherwise, it is discarded. Once the edge effect cuts are applied, we are left with the final galaxy sample. For our SDSS-*GALEX* galaxy sample, $\sim 12\%$ of galaxies are removed from the edge effect cuts. For our PRIMUS galaxy sample, $\sim 40\%$ of galaxies are removed from the edge effect cuts.

In Figure 3.2 we present the distribution of environment measurements (n_{env}) for our final galaxy sample in redshift bins: $z = 0.05 - 0.12$, $0.2 - 0.4$, $0.4 - 0.6$, and $0.6 - 0.8$. The quiescent galaxy contributions are colored in red while the star-forming galaxy contributions are colored in blue and patterned. We classify galaxies with $n_{\mathrm{env}} = 0.0$ to be in low density environments and galaxies with $n_{\mathrm{env}} > 3.0$ to be in high density environments.

Although we imposed PRIMUS redshift errors on our SDSS galaxies to consistently measure environment throughout our entire sample, we note a significant discrepancy between the n_{env} distributions of the SDSS and PRIMUS samples. For example, in each of the PRIMUS redshift bins, $\sim 40\%$ of galaxies in the redshift bin are in low density environments and roughly 30% are in high density environments. In contrast, in the SDSS redshift bin, $\sim 20\%$ of galaxies in the redshift bin are in low density environments and $\sim 35\%$ are in high density environments. We remind the reader that this is mainly due to the varying stellar mass-completeness limits imposed on our galaxy sample for each redshift bins and does not affect our results.

3.4 Results: Stellar Mass Function

Our galaxy sample has so far been classified into quiescent or star-forming and low or high density environments. We further divide these subsamples into redshift bins: $0.05-0.12$, $0.2-0.4$, $0.4-0.6$, and $0.6-0.8$ for a total of 16 subsamples. In Section 3.4.1, we calculate the SMF for each of these subsamples. Then we examine the evolution of active and quiescent subsample SMFs in different environments in Section 3.4.2.

3.4.1 Stellar Mass Function Calculations

To calculate the SMFs we employ a non-parametric $1/V_{\mathrm{max}}$ estimator commonly used for galaxy luminosity functions and stellar mass functions in order to account for Malmquist bias, as done in Moustakas et al. (2013) and discussed in the review Johnston (2011). The differential SMF is given by the following equation:

$$\Phi(\log \mathcal{M})\Delta(\log \mathcal{M}) = \sum_{i=1}^{N} \frac{w_i}{V_{\mathrm{max,avail},i}}. \tag{3.4}$$

w_i is the statistical weight of galaxy i and $\Phi(\log \mathcal{M})\Delta(\log \mathcal{M})$ is the number of galaxies (N) per unit volume within the stellar mass range [$\log\mathcal{M}$, $\log\mathcal{M} + \Delta(\log\mathcal{M})$]. The equation above is the same as Equation 3 in Moustakas et al. (2013) except that we use $V_{\mathrm{max,avail}}$

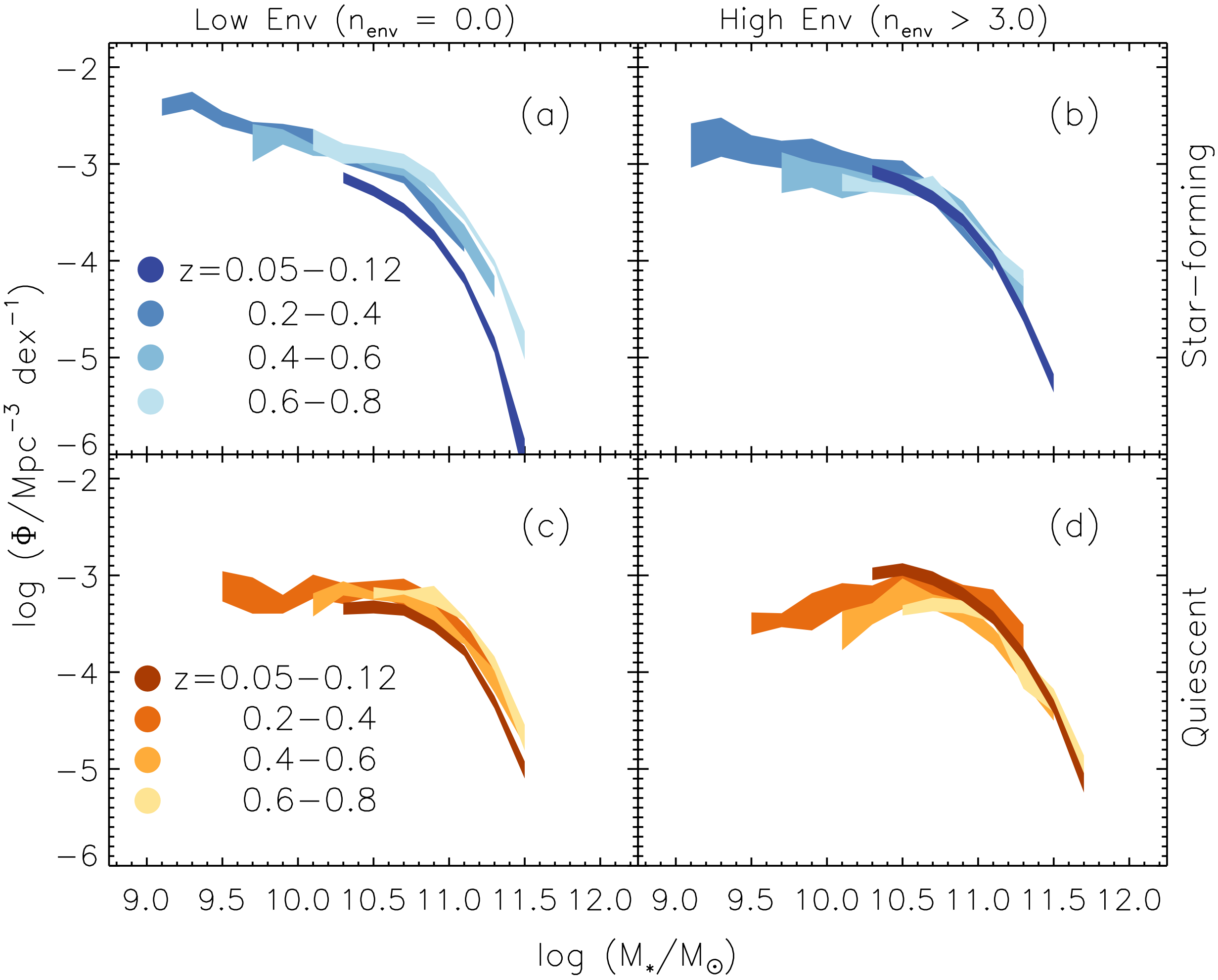

Figure 3.3: Evolution of stellar mass functions of star-forming (top) and quiescent (bottom) galaxies in low (left) and high (right) density environments throughout the redshift range $z = 0$–0.8. The environment of each galaxy was calculated using a cylindrical aperture size of $R = 2.5$ Mpc and $H = 35$ Mpc and classified as low environment when $n_{\rm env} = 0.0$ and as high environment when $n_{\rm env} > 3.0$. The SMFs use mass bins of width $\Delta\log(\mathcal{M}/\mathcal{M}_\odot) = 0.2$. In each panel we use shades of blue (star-forming) and orange (quiescent) to represent the SMF at different redshift, higher redshifts being progressively lighter.

instead than $V_{\rm max}$, to account for the edge effects of the survey discussed in Section 3.3.6.

$V_{{\rm max},i}$ is the maximum cosmological volume where it is possible to observe galaxy i given the apparent magnitude limits of the survey. However in Section 3.3.6 we remove galaxies that lie on the survey edges from our sample. In doing so, we reduce the maximum cosmological volume where a galaxy can be observed, thereby reducing the fraction of $V_{{\rm max},i}$ that is actually available in the sample. We introduce the term $V_{{\rm max,avail},i}$ to express the maximum volume accounting for the survey edge effects.

To calculate $V_{{\rm max,avail},i}$, we use a similar Monte Carlo method as the edge effect cutoffs in Section 3.3.6. First, we generate a sample of points with random RA, Dec within the window function of our galaxy sample (SDSS-*GALEX* window function and the five PRIMUS fields) and random z within the redshift range. These points are not to be confused with the `ransack` sample in Section 3.3.6. We apply the edge effect cuts on these random points as we did for our galaxy sample using the same method as in Section 3.3.6. Within redshift bins of $\Delta z \sim 0.01$, we calculate the fraction of the random points that remain in the bin after the edge effect cuts over the total number of random points in the bin: $f_{\rm edge}$. We then apply this factor to compute $V_{\rm max,avail} = V_{\rm max} \times f_{\rm edge}$. The $V_{\rm max}$ values in the equation above are computed following the method described in Moustakas et al. (2013) Section 4.2 with the same redshift-dependent K-correction from the observed SED and luminosity evolution model.

To calculate the uncertainty of the SMFs from the sample variance, we use a standard jackknife technique (following Moustakas et al. 2013). For the PRIMUS galaxies, we calculate SMFs after excluding one of the five target fields at a time. For the SDSS target galaxies we divide the field into a 12 × 9 rectangular RA and Dec grid and calculate the SMFs after excluding one grid at a time. From the calculated SMFs we calculate the uncertainty:

$$\sigma^j = \sqrt{\frac{N-1}{N}\sum_{k=1}^{M}(\Phi_k^j - \langle\Phi^j\rangle)^2} \tag{3.5}$$

N in this equation is the number of jackknife SMFs in the stellar mass bin. $\langle\Phi^j\rangle$ is the mean number density of galaxies in each stellar mass bin for all of the jackknife Φ^js.

3.4.2 Evolution of the Stellar Mass Function in Different Environments

In Figure 3.3, we present the SMFs of the quiescent/star-forming (orange/blue, bottom/top panels) and high/low density environment (left/right panels) subsamples. The redshift evolution of the SMFs in each of these panels are indicated by a darker shade for lower redshift bins. The width of the SMFs represent the sample variance uncertainties derived in Section 3.4.1.

While a detailed comparison of the SMFs in each panel for different epochs is complicated by the different stellar mass completeness limits, we present some notable trends in each panel. In panel (a), star-forming galaxies in low density environments, we find a significant decrease in the high mass end of the SMF ($\mathcal{M} > 10^{10.75}\mathcal{M}_\odot$) over cosmic time. Meanwhile at lower masses ($\mathcal{M} < 10^{10.5}\mathcal{M}_\odot$), we observe no noticeable trend in the SMF. In panel (b), star-forming galaxies in high density environments, we do not observe any clear trends above the knee of the SMF ($\mathcal{M} \sim 10^{10.7}\mathcal{M}_\odot$) but an increase in SMF below the knee. For the quiescent population in low density environment, panel (c), we observe a potential decrease at higher masses ($\mathcal{M} > 10^{10.7}\mathcal{M}_\odot$). Lastly for the quiescent population in high density environments, panel (d), we find significant increase in Φ for lower masses but little trend at

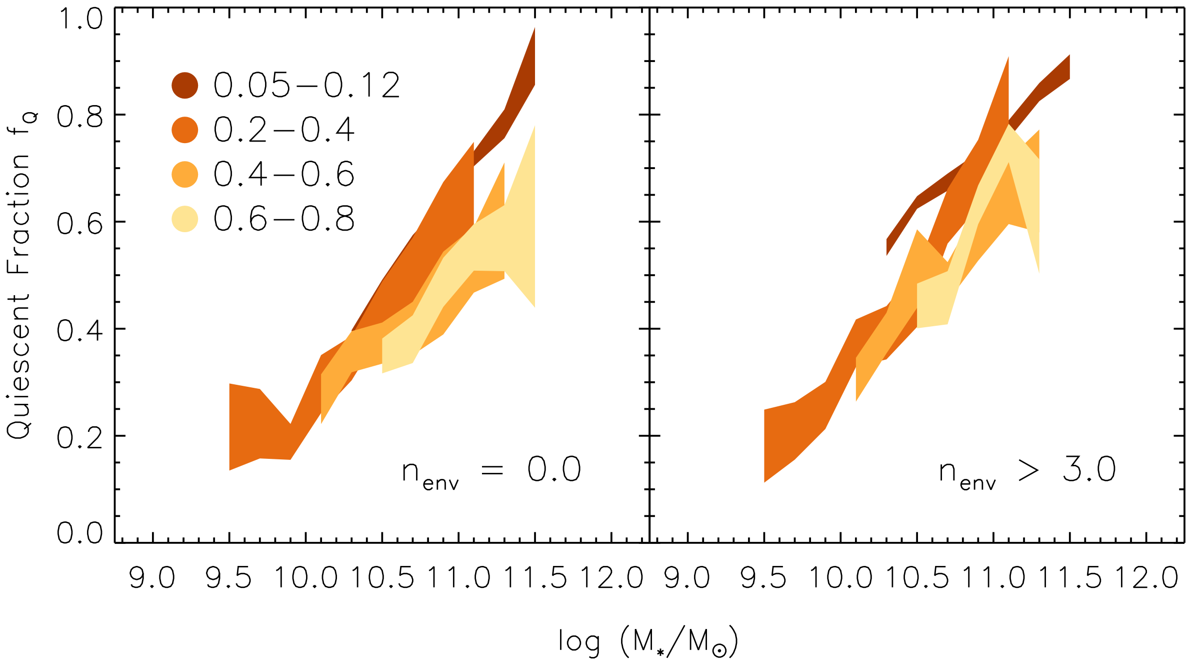

Figure 3.4: Evolution of the quiescent fraction $f_{\rm Q}$ for galaxies in low (left) and high (right) density environments for $z < 0.8$. $f_{\rm Q}$s were calculated using the SMFs in Figure 3.3, as described in Section 3.5.1. Darker shading indicates lower redshift and the width represents the standard jackknife uncertainty.

higher masses.

Observing the evolutionary trends in SMF for each of these sub-populations provides a narrative of the different galaxy evolutionary tracks involving environment and the end of star formation. For example, the decrease in the massive star-forming galaxies in low density environments over cosmic time can be attributed to the transition of those galaxies to any of the other panels. The star-forming galaxies in low density environments that have ended star formation over time are possibly responsible for the increase of the quiescent, low density environment SMF over time. The star-forming galaxies that fall into higher density environments explain the increase in the star-forming high density environment SMF below the knee. Finally, star-forming galaxies in high density environments that have ended their star-formation, quiescent galaxies that have transitioned from low to high density environments, and star-forming galaxies in low density environments that end their star-formation while infalling to high density environments all contribute to the overall increase of the high environment quiescent SMF.

In addition to the evolution over cosmic time, we observe noticeable trends when we compare the SMFs for star-forming and quiescent galaxies between the two environments. Comparison of the SMFs in low versus high density environments reveal a noticeable relation between mass and density, with SMFs in high density environments having more massive galaxies, especially evident in our lowest redshift bin. We further confirm this trend when we compare the median mass between the two environments to find that the median mass for galaxies in high density environments is significantly greater than in low density environments. The relationship between mass and environment observed in our SMFs reflects the well-established mass-density relation and observed mass segregation with environment in the literature (Norberg et al. 2002; Zehavi et al. 2002; Blanton et al. 2005a; Bundy et al. 2006b; Scodeggio et al. 2009; Bolzonella et al. 2010).

While our mass complete subsample coupled with robust environment measurements allows us to compare SMF evolution for each of our subsamples out to $z = 0.8$, we caution readers regarding the photometric biases affecting the SDSS imaging (and perhaps the other imaging sources) and reserve detailed analysis of the SMFs for future investigation.

3.5 Results: Quiescent Fraction

The SMFs calculated in the previous section illustrate the stellar mass distribution of our galaxy population and its evolution over cosmic time. In this section, using the SMFs of our subsamples, we compare the quiescent and the star-forming populations by calculating the fraction of galaxies that have ended their star-formation, the quiescent fraction.

While the fractional relation of the star-forming and quiescent populations has been investigated in the past, with limited statistics, disentangling the environmental effects from underlying correlations among observable galaxy properties such as the color-mass or mass-density relations (Cooper et al. 2010) remains a challenge. With the better statistics available from SDSS and PRIMUS, we evaluate the quiescent fraction in bins of stellar mass, redshift, and environment in Section 3.5.1. By analyzing the quiescent fraction with respect to these properties, in Section 3.5.2 we explicitly compare the quiescent fraction evolution in low and high density environments. Our comparison reveal the subtle environmental effects on the quiescent fraction evolution. Furthermore, by quantifying this environmental effect, we are able constrain the role of environmental effects on how galaxies end their star formation.

3.5.1 Evolution of the Quiescent Fraction

From the SMF number densities (Φ) computed in the previous section, the quiescent fraction is computed as follows,

$$f_{\mathrm{Q}}(\mathcal{M}_*, z) = \frac{\Phi_Q}{\Phi_{SF} + \Phi_Q}. \tag{3.6}$$

Φ_Q and Φ_{SF} are the total number of galaxies per unit volume in stellar mass bin of $\Delta(\log \mathcal{M}) = 0.20$ dex for the quiescent and star-forming subsamples, respectively (Equation 3.4). We compute f_{Q} for high and low density environments for all redshift bins as plotted in Figure 3.4, which shows the evolution of f_{Q} for high (right panel) and low (left panel) density environments. As in Figure 3.3, the evolution of the quiescent fraction over cosmic time is represented in the shading (darker with lower redshift) and the uncertainty is represented by the width. For the uncertainty in the quiescent fraction, we use the standard jackknife technique, following the same steps as for the SMF uncertainty in Section 3.4.1.

Most noticeably in Figure 3.4, we find f_{Q} increases monotonically as a function of mass at all redshifts and environments. In other words, for galaxies in any environment since $z \sim 0.8$, galaxies with higher masses are more likely to have ceased their star-formation. With the roughly linear correlation between galaxy SFR to galaxy color and morphology, we find that this trend reflects the well established color-mass and morphology-mass relations: more massive galaxies are more likely to be red or early-type (Blanton & Moustakas 2009b).

Focusing on the redshift evolution of f_{Q}, we find that for both environments f_{Q} increases as redshift decreases. For high density environments, this is analogous to the Butcher-Oemler Effect (Butcher & Oemler 1984), which states that galaxy populations in groups or clusters have higher f_{blue} (lower f_{Q}) at higher redshift. This evolution occurs with roughly the same amplitude in low environments as well.

In addition, when we compare the stellar masses at which $f_{\mathrm{Q}} = 0.5$ for each subsample, the so-called $\mathcal{M}_{50-50}$, we find that this quantity decreases over cosmic time. This corresponds to the well-known mass-downsizing pattern found by previous investigators (e.g. Bundy et al. 2006b). Furthermore, the mass-downsizing trend observed in each of our environment subsample is qualitatively consistent with the trend observed in zCOSMOS Redshift Survey for isolated and group galaxies (Iovino et al. 2010).

Finally, we compare between our low and high density environment f_{Q}s at each redshift bin interval. For our lowest redshift bin, we find that f_{Q} at low density environments ranges from ~ 0.4 to ~ 0.9 for $10^{10.2}\mathcal{M}_{\odot} < \mathcal{M}_{*} < 10^{11.5}\mathcal{M}_{\odot}$. Over the same mass range, f_{Q} at high density environment ranges from ~ 0.55 to ~ 0.9. For our SDSS sample, f_{Q} in high density environments is notably higher.

For our PRIMUS sample at $z \sim 0.3$, over $10^{9.5}\mathcal{M}_{\odot} < \mathcal{M}_{*} < 10^{11}\mathcal{M}_{\odot}$ f_{Q} ranges from

~ 0.2 to ~ 0.65 for low density environment, while at high density environment $f_{\rm Q}$ ranges from ~ 0.2 to ~ 0.8. Similarly, at $z \sim 0.5$, over $10^{10}\mathcal{M}_\odot < \mathcal{M}_* < 10^{11.2}\mathcal{M}_\odot$ $f_{\rm Q}$ ranges from ~ 0.3 to ~ 0.6 for low density environment and $f_{\rm Q}$ ranges from ~ 0.3 to ~ 0.7 for high density environments. Finally in our highest redshift bin $z \sim 0.7$, over the mass range $10^{10.5}\mathcal{M}_\odot < \mathcal{M}_* < 10^{11.5}\mathcal{M}_\odot$, $f_{\rm Q}$ ranges from ~ 0.35 to ~ 0.6 for low density and ~ 0.45 to ~ 0.8 for high density. For the entire redshift range of our sample, $f_{\rm Q}$ in high density environment is higher than $f_{\rm Q}$ in low density environments.

We note that for $\mathcal{M}_* < 10^{10}\mathcal{M}_\odot$ at $z \sim 0.3$, we find no significant difference between $f_{\rm Q}$ in low and high density environments. Similar quiescent/red fraction studies (e.g. Baldry et al. 2006; Cucciati et al. 2010) find, at these redshifts and mass range, a greater environment dependence in $f_{\rm Q}$. Our classification of star-forming/quiescent galaxies may contribute to this discrepancy with other quiescent fraction studies. We also note that for $\mathcal{M}_* < 10^{10}\mathcal{M}_\odot$ at $z \sim 0.3$, only three of the five PRIMUS fields used in our analysis (XMM-SXDS, XMM-CFHTLS, and COSMOS; see Section 3.3.1) contribute galaxies to our sample. As a result our jack-knife method, which calculates uncertainty by excluding one PRIMUS field at a time, may underestimate the uncertainty thereby making an accurate comparison difficult at low masses. For our analysis, we focus on $\mathcal{M}_* > 10^{10}\mathcal{M}_\odot$.

While there is a significant difference in $f_{\rm Q}$ between the environments, since the difference is observed from our highest redshift bin, it is not necessarily a result of environment dependent mechanisms for ending star formation. In order to isolate any environmental dependence, in the following section we quantitatively compare the evolution of the quiescent fraction between the different environments.

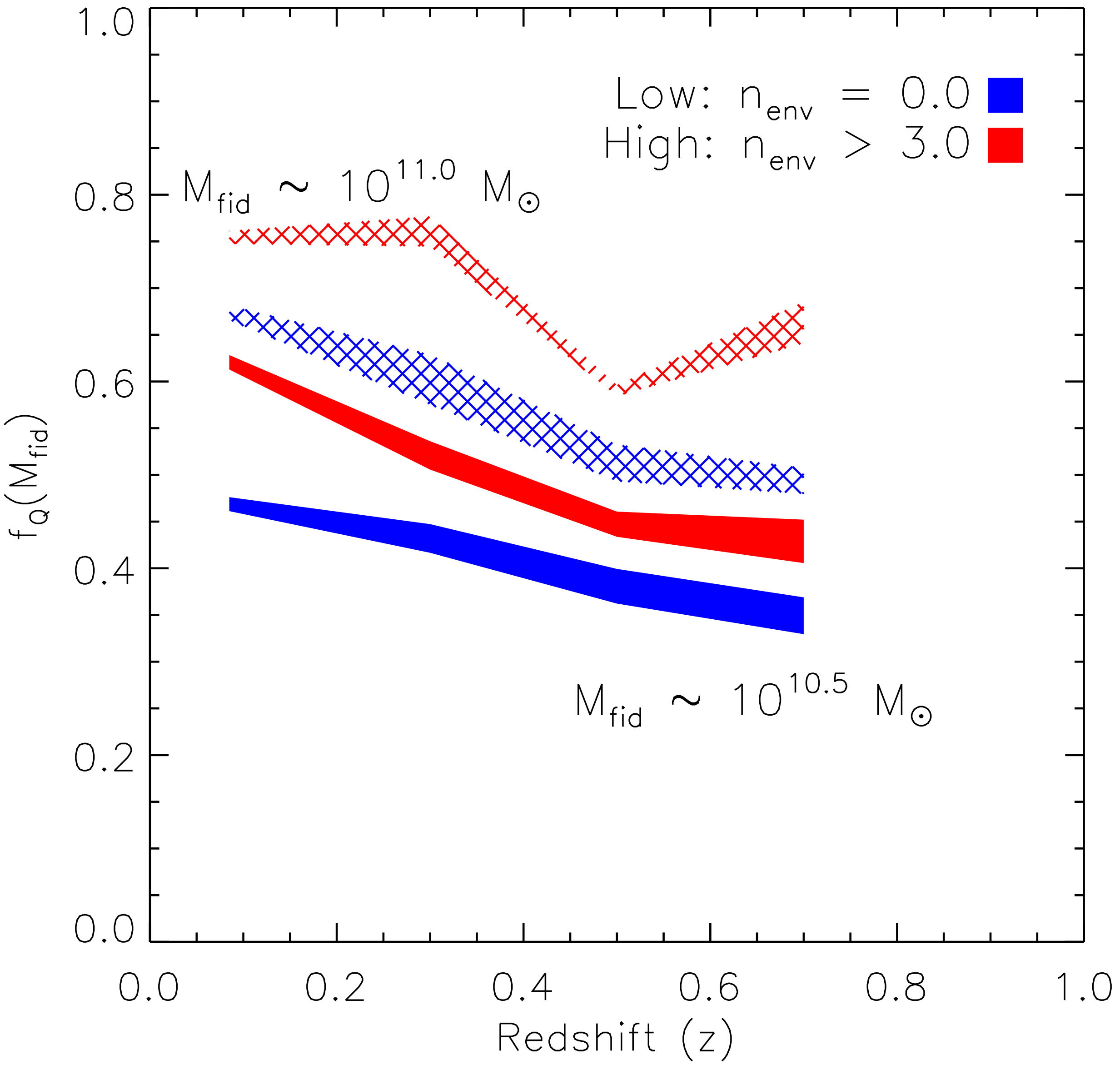

Figure 3.5: The evolution of the quiescent fraction at fiducial mass, $f_Q(\mathcal{M}_{\rm fid})$, for low (blue) and high (red) density environments within the redshift range $z = 0.0 - 0.8$. We present the $f_Q(\mathcal{M}_{\rm fid})$ evolution for $\mathcal{M}_{\rm fid} = 10^{10.5}\mathcal{M}_\odot$ (solid fill) and $10^{11}\mathcal{M}_\odot$ (patterned fill) with the uncertainty of the best-fit parameter b in Equation 3.7 represented by the width of the line. While the high density $f_{\rm Q}(\mathcal{M}_{\rm fid})$ is greater than low density environment $f_{\rm Q}(\mathcal{M}_{\rm fid})$ over the entire redshift range of our sample, there is a significant increase in $f_{\rm Q}(\mathcal{M}_{\rm fid})$ over cosmic time for both environments. For the environment cut-offs ($n_{\rm env} = 0.0$ for low and $n_{\rm env} > 3.0$ for high), there is no significant difference in the slope of the evolution between the environments.

Table 3.2: Best Fit Parameters for $f_{\rm Q}(\mathcal{M}_*)$ Fit

$z_1 < z < z_2$	Environment	a	b
$0.05 < z < 0.12$	$n_{\rm env} = 0.0$	0.410 ± 0.018	0.469 ± 0.007
	$n_{\rm env} > 3.0$	0.270 ± 0.016	0.620 ± 0.008
$0.2 < z < 0.4$	$n_{\rm env} = 0.0$	0.340 ± 0.032	0.432 ± 0.015
	$n_{\rm env} > 3.0$	0.432 ± 0.018	0.544 ± 0.010
$0.4 < z < 0.6$	$n_{\rm env} = 0.0$	0.263 ± 0.038	0.381 ± 0.018
	$n_{\rm env} > 3.0$	0.289 ± 0.018	0.446 ± 0.013
$0.6 < z < 0.8$	$n_{\rm env} = 0.0$	0.284 ± 0.036	0.352 ± 0.019
	$n_{\rm env} > 3.0$	0.468 ± 0.065	0.429 ± 0.023

Notes: Best fit parameters in Equation 3.7 for each subsample $f_{\rm Q}(\mathcal{M}_*)$ in Figure 3.4 for $\mathcal{M}_{\rm fid} = 10^{10.5}\mathcal{M}_\odot$.

3.5.2 Environmental Effects on the Quiescent Fraction Evolution

In order to more quantitatively compare the $f_{\rm Q}$ evolution for different epochs and environments, we fit $f_{\rm Q}$ for each subsample to a power-law parameterization as a function of stellar mass,

$$f_{\rm Q}(\mathcal{M}_*) = a \log\left(\frac{\mathcal{M}_*}{\mathcal{M}_{\rm fid}}\right) + b, \tag{3.7}$$

where a and b are best-fit parameters using *MPFIT* (Markwardt 2009) and $\mathcal{M}_{\rm fid}$ represents the empirically selected fiducial mass within the stellar mass limits where there is a sufficiently large number of galaxies. We primarily focus on $\mathcal{M}_{\rm fid} = 10^{10.5}\,\mathcal{M}_\odot$.

In Figure 3.5 we present the evolution of $f_{\rm Q}(\mathcal{M}_{\rm fid})$ from $z \sim 0.7$ to ~ 0.1 at low (blue) and high (red) density environments for $\mathcal{M}_{\rm fid} = 10^{10.5}\,\mathcal{M}_\odot$ (solid fill) and $10^{11}\,\mathcal{M}_\odot$ (pattern fill). The width of the evolution represents the uncertainty derived from *MPFIT*. As noted earlier in Section 3.5.1, $f_{\rm Q}$ in high density environments is significantly greater than $f_{\rm Q}$ in low density environments for both fiducial mass choices. Throughout our sample's redshift range $f_{\rm Q}(\mathcal{M}_{\rm fid})_{\rm high} - f_{\rm Q}(\mathcal{M}_{\rm fid})_{\rm low} \sim 0.1$.

In addition, the $f_{\rm Q}(\mathcal{M}_{\rm fid})$ evolution illustrates that the quiescent fraction in low density environment increases over cosmic time: $f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.1) - f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.7) \sim 0.1$. This significant quiescent fraction evolution for low density environments suggests that internal mechanisms, independent of environment, are responsible for a significant amount of star-formation cessation. Meanwhile, the $f_{\rm Q}(\mathcal{M}_{\rm fid})$ evolution in high density environment ($f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.1) - f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.7) \sim 0.12$) shows little additional evolution.

When we increase our choice of $\mathcal{M}_{\rm fid}$ to $10^{11}\mathcal{M}_\odot$, aside from an overall shift in $f_{\rm Q}(\mathcal{M}_{\rm fid})$ by ~ 0.2, we observe the same evolutionary trends. $f_{\rm Q}(\mathcal{M}_{\rm fid} = 10^{11}\mathcal{M}_\odot)$ for both low and high density environments each increase by ~ 0.2 from at all redshifts we study. Increasing the fiducial mass to $10^{11}\mathcal{M}_\odot$ does not significantly alter the evolutionary trends in either environment. Although the varying stellar mass completeness at each redshift bin limits the masses we probe for the $f_{\rm Q}$ evolution, our $f_{\rm Q}$ evolution exhibits little mass dependence.

However, the uncertainties in the PRIMUS redshifts may contaminate our fixed aperture measurements of galaxy environment. Consequently, we consider in Figure 3.6 more stringent high density environment classifications, extending the cut off to $n_{\rm env} > 5$ and 7 (specified in the top right legend and represented by the color of the shading). Aside from the increase in uncertainties that accompany the decrease in sample size of the purer high environment sample, we find an extension of the $f_{\rm Q}$ difference between the environments we stated earlier. A more stringent high environment classification significantly increases the overall $f_{\rm Q}(\mathcal{M}_{\rm fid})$, which rises monotonically with the $n_{\rm env}$ limit.

More importantly, a purer high environment classification reveals a more significant environment dependence on the $f_{\rm Q}$ evolution. While the difference between the $f_{\rm Q}$ evolution in low and high density environment is negligible for the $n_{\rm env} > 3$ cut-off, there is a notable difference in $f_{\rm Q}$ evolution between our highest cut-off $n_{\rm env} > 7$ and our low density environment. $f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.1) - f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.7) \sim 0.25$ for $n_{\rm env} > 7$ versus $f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.1) - f_{\rm Q}(\mathcal{M}_{\rm fid}, z \sim 0.7) \sim 0.1$ for low density environment. In addition to the environment independent internal mechanisms that can explain the $f_{\rm Q}$ evolution in low density environments, there may be other environment dependent mechanisms that can account for the moderate environment dependence of the $f_{\rm Q}$ evolution. Our measured difference in the $f_{\rm Q}$ evolution between environments provides an important constraint for any environmental models for ending star formation.

3.5.3 Comparison to Literature

Although a direct comparison with other results is difficult due to our sample specific methodology, a number of results from the literature have investigated the quiescent fraction in comparable fashions. In this section we compare our $f_{\rm Q}$ results from above to a number of these results, specifically from SDSS and zCOSMOS, with similarly defined samples and analogous environment classifications.

In Figure 3.6, we plot best-fit parameterization of $f_{\rm red}$ for high and low density environment from SDSS (panel a), zCOSMOS (panel b), and Peng et al. (2010) (filled square; panel d) from both surveys. From Iovino et al. (2010) (empty square; panel b), we calculate $f_{\rm red} = 1 - f_{\rm blue}$ using the best-fit $f_{\rm blue}$ from the mass bin $\mathcal{M} = 10^{10.3} - 10^{10.8}\mathcal{M}_\odot$. From Kovač et al. (2014) (triangle; panel b) we plot an estimated $f_{\rm Q}$ by applying the residual between SFR based and color based galaxy classifications to the best-fit $f_{\rm red}$ at $\mathcal{M} = 10^{10.5}\mathcal{M}_\odot$ for low ($\delta = 0.0$) and high density environments ($\delta = 1.5$). Similarly, from Baldry et al. (2006) (diamond; panel a) we plot $f_{\rm Q}$ derived from the best-fit $f_{\rm red}$ at $\mathcal{M} = 10^{10.5}\mathcal{M}_\odot$ for low ($\delta = 0.0$) and high density environment ($\delta = 1.0$). For Geha et al. (2012a) (cross; panel a), we plot $f_{\rm Q}$ for their isolated galaxy sample in their mass bin closest to $10^{10.5}\mathcal{M}_\odot$,

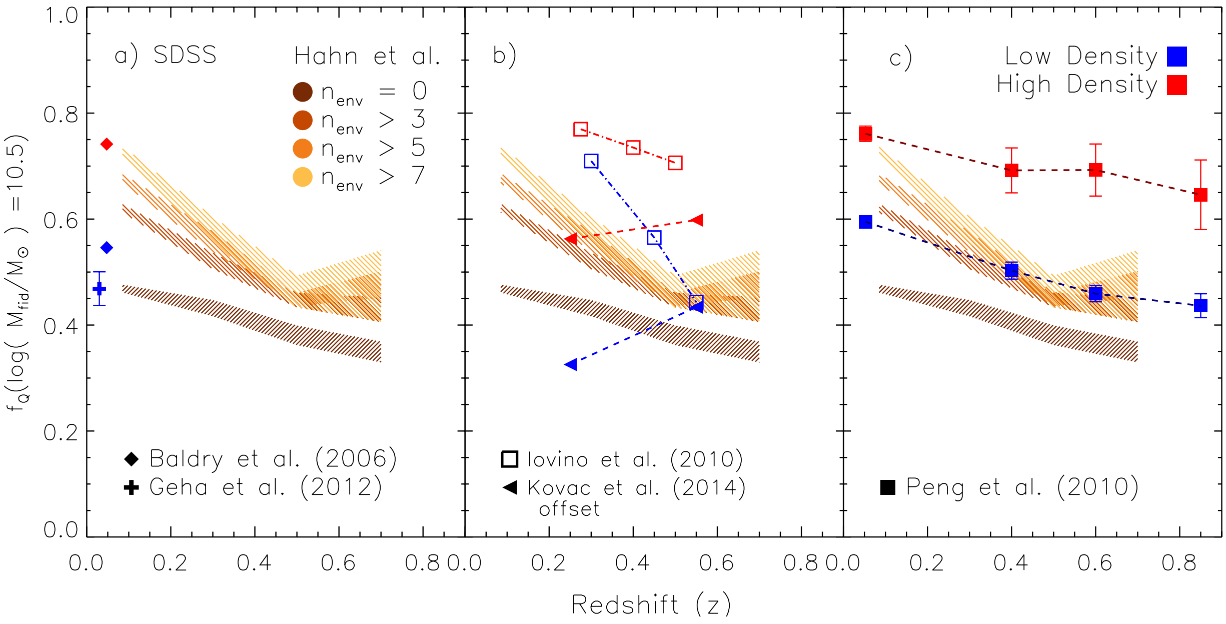

Figure 3.6: $f_Q(\mathcal{M}_{fid} = 10^{10.5}\mathcal{M}_\odot)$ evolution compared to $f_{red}(\mathcal{M}_* \sim 10^{10.5}\ \mathcal{M}_\odot)$ in the literature: Baldry et al. (2006) (diamond) and Geha et al. (2012a) (cross) from SDSS (panel a), Iovino et al. (2010) (empty square) and Kovač et al. (2014) (triangle) from zCOSMOS (panel b), and Peng et al. (2010) from both SDSS and zCOSMOS (panel c). The f_{red} values from Iovino et al. (2010), Kovač et al. (2014), Baldry et al. (2006), and Peng et al. (2010) are calculated from the best-fit parameterizations presented in the respective works. High density environment is represented in red and low density environment is represented in blue. The f_Q value from Geha et al. (2012a) is the f_Q value at $\mathcal{M} = 10^{10.55}\mathcal{M}_\odot$. Uncertainties in the Iovino et al. (2010) best-fit f_{red} is omitted due to insufficient information on the cross correlation terms of the fit parameters. For Kovač et al. (2014) we apply the offset between the color-based and SFR-based galaxy classification in order to plot the f_Q estimates. We also plot the $f_Q(\mathcal{M}_{fid} = 10^{10.5}\mathcal{M}_\odot)$ evolution of our sample with varying environment cut-offs specified on the top right. As in Figure 3.5 the width of the $f_Q(\mathcal{M}_{fid} = 10^{10.5}\mathcal{M}_\odot)$ evolution represent the uncertainty in the best-fit parameters of Equation 3.7.

$\mathcal{M} = 10^{10.55}\mathcal{M}_{\odot}$. Finally for Peng et al. (2010) (square; panel c), we plot the parameterized f_{red} at $\mathcal{M} = 10^{10.5}\mathcal{M}_{\odot}$ using their best-fit parameters for low ($\delta = 0.0$) and high ($\delta = 1.4$) density environments.

For our lowest redshift bin SDSS sample, we find that our f_{Q} for low and high environments are consistent with other SDSS f_{Q} (or f_{red}) measurements as a function of environment. For example, Baldry et al. (2006) uses projected neighbor density environment measures ($\log\Sigma$) to obtain $f_{\mathrm{Q}}(\mathcal{M})$ for a range of environmental densities. Although the different environment measurements make direct comparisons difficult, in their corresponding higher environments ($\log\Sigma > 0.2$ in Baldry et al. 2006) $f_{\mathrm{Q}}(\mathcal{M} \sim 10^{10.2}\mathcal{M}_{\odot}) \sim 0.6$ and $f_{\mathrm{Q}}(\mathcal{M} \sim 10^{11.5}\mathcal{M}_{\odot}) \sim 0.9$, which is in agreement with our high density environment. Likewise, for lower environments ($\log\Sigma < -0.4$ in Baldry et al. 2006) $f_{\mathrm{Q}}(\mathcal{M} \sim 10^{10.2}\mathcal{M}_{\odot}) \sim 0.4$ and $f_{\mathrm{Q}}(\mathcal{M} \sim 10^{11.5}\mathcal{M}_{\odot}) \sim 0.8$, which also agree with our low density environment f_{Q}. The Baldry et al. (2006) points (diamond) in Figure 3.6 reflect this agreement.

More recently, Tinker et al. (2011), using a group-finding algorithm on the SDSS DR7, presents the relationship between f_{Q} and overdensity for galaxies within the mass range $\log\mathcal{M} = [9.8, 10.1]$. The Tinker et al. (2011) f_{Q} at the lowest and highest overdensities, $f_{\mathrm{Q}} \sim 0.4$ and $f_{\mathrm{Q}} \sim 0.6$ respectively, are consistent with our f_{Q} for low and high density environment at the lower mass limit ($\log\mathcal{M} \sim 10.2$).

A modified Tinker et al. (2011) sample is used in Geha et al. (2012a) to obtain f_{Q} for isolated galaxies over a wider mass range ($10^{7.4}\mathcal{M}_{\odot}$ to $10^{11.2}\mathcal{M}_{\odot}$). Although Geha et al. (2012a) probe a slightly lower redshift range ($z \leq 0.06$), their f_{Q} is consistent with our low density sample. Within the overlapping mass range, at the low mass end Geha et al. (2012a) find $f_{\mathrm{Q}}(\mathcal{M}_{*} \sim 10^{10.2}\mathcal{M}_{\odot}) \sim 0.3$ and at the high mass end they find $f_{\mathrm{Q}}(\mathcal{M}_{*} \sim 10^{11.2}\mathcal{M}_{\odot}) \sim 0.8$. Both of these values agree with our lowest redshift f_{Q} results in low density environment. Figure 3.6 illustrates the f_{Q} agreement for $\mathcal{M}_{*} = 10^{10.5}\mathcal{M}_{\odot}$.

For $z > 0.2$, we compare our PRIMUS $f_{\rm Q}$ results to the $f_{\rm red}$ (or $1 - f_{\rm blue}$) results from the zCOSMOS Redshift Survey (Iovino et al. 2010; Kovač et al. 2014), which covers a similar redshift range as PRIMUS. Iovino et al. (2010), and Kovač et al. (2014) using a mass-complete galaxy sample derived from zCOSMOS and a group catalog, 3D local density contrast, and overdensity environment measurements, respectively, compare $f_{\rm red}$ with respect to environment. The $f_{\rm blue}$ for group and isolated galaxies from Iovino et al. (2010) are generally inconsistent with our $1 - f_{\rm Q}$ for high and low density environments.

Similarly, $f_{\rm red}$ for high and low overdensities in Kovač et al. (2014) are greater overall than the PRIMUS $f_{\rm Q}$ values in high and low density environments. However, Kovač et al. (2014) points out that there is a significant difference between classifying the quiescent population using color and SFR due to dust-reddening in star-forming galaxies. For their lower redshift bin ($0.1 < z < 0.4$) Kovač et al. (2014) find that their $f_{\rm Q}$ defined by color is greater than $f_{\rm Q}$ defined by SFR by roughly 0.2. While for their higher redshift bin ($0.4 < z < 0.7$) the difference is $0.15 - 0.19$. Although Kovač et al. (2014) does not elaborate on how the galaxy classification discrepancy applies to the different environments, if we simply account for the difference uniformly for $f_{\rm red}$ at all environments, the Kovač et al. (2014) results in their lower redshift bin are roughly consistent with our $f_{\rm Q}$ at high and low density environments. Even accounting for the dust-reddening of $f_{\rm red}$, Kovač et al. (2014) finds a significantly higher $f_{\rm Q}$ in their higher redshift bin.

In Figure 3.4, the $f_{\rm Q}$ evolution with respect to mass reveals, qualitatively, little mass dependence in the evolution. Moreover, in Figure 3.5, we illustrated that adjusting the fiducial mass only shifted the overall $f_{\rm Q}(\mathcal{M}_{\rm fid})$, but did not change the $f_{\rm Q}$ evolutionary trend. The consistency in the $f_{\rm Q}$ evolutionary trends over change in fiducial mass suggests that $f_{\rm Q}$ evolution exhibit little mass dependence within the mass range probed in our analysis. In contrast to the weak mass dependence we observe in our results, Iovino et al. (2010) find significantly different $f_{\rm Q}$ evolution at $\mathcal{M} \sim 10^{11}\mathcal{M}_\odot$ and $\mathcal{M} \sim 10^{10.5}\mathcal{M}_\odot$, for both group and isolated galaxies. In fact at their highest mass bin ($10^{10.9} - 10^{11.4}\mathcal{M}_\odot$), Iovino et al. (2010) find no evolution for both environments: constant $f_{\rm blue} \sim 0.1$ over $z = 0.3 - 0.8$ for both group and isolated galaxy populations.

Meanwhile in their mass bin most comparable to $\mathcal{M}_{\rm fid} \sim 10^{10.5}\mathcal{M}_\odot$ ($10^{10.3}\mathcal{M}_\odot - 10^{10.8}\mathcal{M}_\odot$), Iovino et al. (2010) finds that $f_{\rm blue}$ evolves by ~ 0.1 from $z = 0.5$ to 0.25 for group galaxies and by ~ 0.3 from $z = 0.55$ to 0.3 for isolated galaxies as presented in panel (b) of Figure 3.6. Altogether, with mass bins beyond the fiducial masses we explore, Iovino et al. (2010) find a strong mass dependence with $f_{\rm Q}$ evolving significantly more in lower mass bins. While our sample from PRIMUS provides larger statistics than zCOSMOS, the mass-completeness limits we impose on our sample limits the mass range we probe (e.g. $\mathcal{M} > 10^{10.5}\mathcal{M}_\odot$ for our $z \sim 0.7$ bin). Consequently our results cannot rule out mass dependence in the $f_{\rm Q}$ evolution at lower masses.

In Figure 3.5 and Figure 3.6 we quantified that throughout our redshift range, high density environments have a significantly greater $f_{\rm Q}(\mathcal{M}_{\rm fid})$ than the low density environments. This finding is in agreement with the zCOSMOS results from Cucciati et al. (2010) and Kovač et al. (2014). As illustrated in panel (b) of Figure 3.6, Kovač et al. (2014) finds $f_{\rm Q}$ in high density environment significantly greater than $f_{\rm Q}$ at low density environment. Moreover, since galaxy color serves as a proxy for SFR, our results support the existence of the color-density relation (Cucciati et al. 2010; Cooper et al. 2010) and is not consistent with the color-density relation being merely a reflection of the mass-density relationship, as Scodeggio et al. (2009) suggest it is based on the Vimos VLT Deep Survey ($0.2 < z < 1.4$).

In Section 3.5.2, we showed that $f_{\rm Q}$ in low density environments evolves over cosmic time. From this trend we deduce that internal, environment independent, mechanisms contribute to ending star-formation in galaxy evolution. Iovino et al. (2010) from zCOSMOS, plotted in

Figure 3.6 panel (b), also find that $f_{\rm Q}$ in low density environment increases with decreasing redshift. On the other hand Kovač et al. (2014), also from zCOSMOS, presents that $f_{\rm Q}$ in low density environment decreases over cosmic time. While the uncertainties for the parameterized $f_{\rm Q}$ are not listed, and thus not shown in Figure 3.6, once they are accounted for, Kovač et al. (2014) find no significant $f_{\rm Q}$ evolution over cosmic time. However, once we account for the dust-reddening of the $f_{\rm red}$, we find a more significant decrease over cosmic time (Figure 3.6 panel b).

Furthermore, in Section 3.5.2, our comparison of the $f_{\rm Q}$ evolution between the lowest density environment and the highest density environment revealed a modicum of evidence for the existence of environment dependent mechanisms. The same comparison with zCOSMOS results (Iovino et al. 2010; Kovač et al. 2014) present trends inconsistent with our findings. First, comparing the high (red) and low (blue) density environments for Iovino et al. (2010) in Figure 3.6 shows that there are indeed pronounced discrepancies between the $f_{\rm Q}$ evolution in different environments. Group galaxies in Iovino et al. (2010) have higher overall $f_{\rm Q}$ than isolated galaxies. However, unlike our results, which find a greater $f_{\rm Q}$ evolution at higher density environments, $f_{\rm Q}$ in Iovino et al. (2010) shows the opposite environment dependence that there is a significantly greater $f_{\rm Q}$ evolution for isolated galaxies. Once the large uncertainties in the $f_{\rm Q}$ fit are taken into account, Iovino et al. (2010) state that the $f_{\rm Q}$ is difficult to measure from their sample.

Next, Kovač et al. (2014) also find that overall $f_{\rm Q}$ is greater in high density than in low density environments. Like their low density environment $f_{\rm Q}$ evolution, $f_{\rm Q}$ in high density environment decreases over cosmic time between their two redshift bins. Although the decrease in $f_{\rm Q}$ over cosmic time conflicts with our results, Kovač et al. (2014) finds a greater (less negative) $f_{\rm Q}$ evolution in high density environments relative to low density environments, suggesting an environment dependence that is in the same direction as our results. We note that the negative slopes of the f_{Q} evolution in both environments are enhanced in Figure 3.6 due to the dust-reddening correction we impose to the Kovač et al. (2014) f_{red} results.

Due to the redshift uncertainties in PRIMUS, our galaxy environment measures are more susceptible to contamination. As discussed in Coil et al. (2011) and Cool et al. (2013), PRIMUS has redshift success rate of $> 75\%$; in comparison, zCOSMOS has 88% redshift completeness for the entire sample and 95% complete within the redshift range $0.5 < z < 0.8$ (Lilly et al. 2009). Although the zCOSMOS survey provides more precise spectroscopic redshifts, PRIMUS has higher overall completeness due to its high targeting fraction of $\sim 80\%$. zCOSMOS has a spatial sampling rate of $\sim 30 - 50\%$ and a overall completeness rate of $48 - 52\%$ (Knobel et al. 2012). Our sample also provides larger statistics and covers a larger portion of the sky. Our SDSS-*GALEX* sample covers $2,505$ deg^2. More comparably, our PRIMUS sample covers 5.5 deg^2, over 3 times the sky coverage of zCOSMOS (1.7 deg^2). Furthermore, our PRIMUS sample is constructed from five independent fields which allows us to significantly reduce the effects of cosmic variance.

As listed in Table 3.1, after our edge effect cuts and stellar mass completeness limits, our sample consists of $13,734$ galaxies from PRIMUS over $0.2 < z < 0.8$ and $63,417$ galaxies from SDSS over $0.05 < z < 0.12$. Meanwhile, Iovino et al. (2010) has 914 galaxies with $\mathcal{M} > 10^{10.3}\mathcal{M}_\odot$ over $0.1 < z < 0.6$ and 1033 galaxies with $\mathcal{M} > 10^{10.6}\mathcal{M}_\odot$ over $0.1 < z < 0.8$. For the actual sample used to obtain the best-fit f_{Q} values in Figure 3.6 Iovino et al. (2010) has 617 galaxies. In comparison, our PRIMUS sample alone contains > 20 times the number of galaxies. While there is a considerable difference in the overall f_{Q} between our results and those of Iovino et al. (2010), the use of different methodologies, particularly for galaxy classification and environment measurements, make such comparisons ambiguous. On the other hand, the discrepancies in the f_{Q} evolutionary trends with our results may be explained by the limited statistics in the Iovino et al. (2010) sample.

The more recent Kovač et al. (2014) provides larger statistics with $2,340$ galaxies in their lower redshift bin ($0.1 < z < 0.4$) and $2,448$ galaxies in their higher redshift bin ($0.4 < z < 0.7$). Although their sample is smaller than the PRIMUS sample, which contains over twice times the number of galaxies, the Kovač et al. (2014) sample provides a more stable comparison. Once their results are adjusted for the dust-reddening, we find that their overall $f_{\rm Q}$ is more or less consistent with our overall $f_{\rm Q}$. However, it is difficult to explain the significant discrepancies in the $f_{\rm Q}$ evolutionary trends. The significant overdenities observed in the COSMOS field at $z \sim 0.35$ and $z \sim 0.7$ (Lilly et al. 2009; Kovač et al. 2010) may have a significant effect on the zCOSMOS results and offer a possible explanation for the discrepancies.

3.6 Summary and Discussion

Using a stellar mass complete galaxy sample derived from SDSS and PRIMUS accompanied by a consistently measured galaxy environment from robust spectroscopic redshifts, we measure the stellar mass functions for star-forming and quiescent galaxies in low and high density environments over the redshift range $0.05 < z < 0.8$. From these stellar mass functions, we compare the proportion of galaxies that have ended their star-formation within the subsamples by computing the quiescent fraction for each of them. In order to better quantify the evolution of the quiescent fraction over cosmic time, we fit our quiescent fraction anchored at a fiducial mass.

From our analyses we find the following notable results. The first three demonstrate that previous findings that are well known in the local universe are applicable out to $z \sim 0.7$. The last two are consistent with the findings of Peng et al. (2010) but provide increased detail on the environmental dependence of galaxy evolution:

1. From the SMFs, we find that the galaxy population in high density environments, both star-forming and quiescent, have a higher median mass, thus confirming the mass-density relation and mass-segregation in different environments throughout our sample's redshift range.

2. For all subsamples, $f_{\rm Q}$ increases monotonically with galaxy stellar mass, showing a clear mass dependence and reflecting the well-established color-mass and morphology-mass relations.

3. We illustrate that $f_{\rm Q}$ in high density environments is greater than $f_{\rm Q}$ in low density environments for $\mathcal{M} \sim 10^{10.5} - 10^{11} \mathcal{M}_\odot$ and out to redshift $z \sim 0.7$. This result reflects the well known trend that galaxies in high density environment are statistically redder, have lower SFRs, and are more massive.

4. $f_{\rm Q}$ increases significantly with redshift for both low and high density environments. For high density environment, this trend is the Butcher-Oemler effect. Furthermore, the $f_{\rm Q}$ evolution in low density environment suggest the existence of internal environment-independent mechanisms for ending star formation.

5. Comparison of the $f_{\rm Q}(\mathcal{M}_{\rm fid})$ evolution for a range of environment classifications reveals that the since $z = 0.8$, $f_{\rm Q}$ has evolved by a greater amount in the highest density environments. For our purest high environment sample ($n_{\rm env} > 7$), the total $f_{\rm Q}$ evolution is ~ 0.1 greater than the total $f_{\rm Q}$ evolution in low density environment, revealing a moderate dependence on environment.

Many physical mechanisms have been proposed to explain the cessation of star-formation observed in many galaxies. Recently star-formation cessation has often been classified into internal or external mechanisms, and sometimes more specifically into mass-dependent and environment-dependent mechanisms (Baldry et al. 2006; Peng et al. 2010). The significant redshift evolution of the $f_{\rm Q}$ in low density environments confirms the existence of internal mechanisms that end star-forming in galaxies.

Furthermore, the greater $f_{\rm Q}$ evolution in the highest density environment relative to low density environments suggests that in addition to the internal mechanisms, in high density environments such as groups and clusters, environment-dependent effects may also contribute to the end of star-formation. Our results do not specifically shed light on which mechanisms (e.g. strangulation, ram-pressure stripping, etc.) occur in high density environments. Not to mention, the mechanism could yet be indirect; for example, the galaxies in higher density environments could end star-formation primarily due internal processes, which affect the galaxies that end up in groups and clusters more greatly. Nevertheless, our results impose important constraints on the total possible contribution of environment dependent mechanisms that models must satisfy, providing a limit on the role of environment in ending star formation in galaxies.

Acknowledgements

CH and MB were supported by NSF-AST-1109432 and NSF-AST-0908354. MB was also supported by NSF-AST-1211644. ALC acknowledges support from NSF CAREER award. We acknowledge Katarina Kovač and Ying-Jie Peng for insightful discussions. We also thank Marla Geha for providing published quiescent fraction results in electronic format.

Funding for PRIMUS has been provided by NSF grants AST-0607701, 0908246, 0908442, 0908354, and NASA grant 08-ADP08-0019. The Galaxy Evolution Explorer (GALEX) is a NASA Small Explorer, launched in April 2003, whose mission was developed in cooperation with the Centre National d‘Etudes Spatiales of France and the Korean Ministry of Science and Technology. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions.

Chapter 4

Star Formation Quenching Timescale of Central Galaxies in a Hierarchical Universe

This Chapter is joint work with Jeremy L. Tinker (NYU) and Andrew Wetzel (UC Davis) submitted to *The Astrophysical Journal* as Hahn et al. (2017).

4.1 Chapter Abstract

Central galaxies make up the majority of the galaxy population, including the majority of the quiescent population at $\mathcal{M}_* > 10^{10}\mathrm{M}_\odot$. Thus, the mechanism(s) responsible for quenching central galaxies plays a crucial role in galaxy evolution as whole. We combine a high resolution cosmological N-body simulation with observed evolutionary trends of the "star formation main sequence," quiescent fraction, and stellar mass function at $z < 1$ to construct a model that statistically tracks the star formation histories and quenching of central galaxies. Comparing this model to the distribution of central galaxy star formation rates in a group catalog of the SDSS Data Release 7, we constrain the timescales over which physical processes cease star formation in central galaxies. Over the stellar mass range $10^{9.5}$ to $10^{11}\mathrm{M}_\odot$ we infer quenching e-folding times that span 1.5 to 0.5 Gyr with more massive central galaxies quenching faster. For $\mathcal{M}_* = 10^{10.5}\mathrm{M}_\odot$, this implies a total migration time of ~ 4 Gyrs from the star formation main sequence to quiescence. Compared to satellites, central galaxies take ~ 2 Gyrs longer to quench their star formation, suggesting that different mechanisms are responsible for quenching centrals versus satellites. Finally, the central galaxy quenching timescale we infer provides key constraints for proposed star formation quenching mechanisms. Our timescale is generally consistent with gas depletion timescales predicted by quenching through strangulation. However, the exact physical mechanism(s) responsible for this still remain unclear.

4.2 Introduction

Observations of galaxies using large galaxy surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), Cosmic Evolution Survey (COSMOS; Scoville et al. 2007), and the PRIsm MUlti-object Survey (PRIMUS; Coil et al. 2011; Cool et al. 2013) have firmly established a global view of galaxy properties out to $z \sim 1$. Galaxies are broadly divided into two main classes: star forming and quiescent. Star forming galaxies are blue in color, forming stars, and typically disk-like in morphology. Meanwhile quiescent galaxies are red in color, have little to no star formation, and typically have elliptical morphologies (Kauffmann et al. 2003; Blanton et al. 2003; Baldry et al. 2006; Wyder et al. 2007; Moustakas et al. 2013; for a recent review see Blanton & Moustakas 2009a).

Over the period $z < 1$, detailed observations of the stellar mass functions (SMF) reveal a significant decline in the number density of massive star forming galaxies accompanied by an increase in the number density of quiescent galaxies (Blanton et al. 2006; Borch et al. 2006b; Bundy et al. 2006a; Moustakas et al. 2013). The growth of the quiescent fraction with cosmic time also reflects this change in galaxy population (Peng et al. 2010; Tinker et al. 2013; Hahn et al. 2015). Imprints of galaxy environment on the quiescent fraction (Hubble 1936b; Oemler 1974; Dressler 1980; Hermit et al. 1996; for a recent review see Blanton & Moustakas 2009a) suggest that there is a significant correlation between environment and the cessation of star formation. In comparison to the field, high density environments have a higher quiescent fraction. However, observations find quiescent galaxies in the field (Baldry et al. 2006; Tinker et al. 2011; Geha et al. 2012b), at least for galaxies with stellar mass down to $10^9 \mathrm{M}_\odot$ (Geha et al. 2012b), and as Hahn et al. (2015) finds using PRIMUS, the quiescent fraction in both high density environments and the field increase significantly over time.

Furthermore, galaxy environment is a subjective and heterogeneously defined quantity in the literature (Muldrew et al. 2012). It can, however, be more objectively determined within the halo occupation context, which labels galaxies as 'centrals' and 'satellites' (Zheng et al. 2005b; Weinmann et al. 2006; Blanton & Berlind 2007; Tinker et al. 2011). Central galaxies reside at the core of their host halos while satellite galaxies orbit around. During their infall, satellite galaxies are likely to experience environmentally driven mechanisms such as ram pressure stripping (Gunn & Gott 1972; Bekki 2009), strangulation (Larson et al. 1980; Balogh et al. 2000), or harassment (Moore et al. 1998).

Central galaxies, within this context, are thought to cease their star formation through internal processes – numerous mechanisms have been proposed and demonstrated on semi-analytic models (SAMs) and hydrodynamic simulations. One common proposal explains that hot gaseous coronae form in halos with masses above $\sim 10^{12} \mathrm{M}_\odot$ via virial shocks, which starve galaxies of cool gas required to fuel star formation (Birnboim & Dekel 2003; Kereš et al. 2005; Croton et al. 2006; Cattaneo et al. 2006; Dekel & Birnboim 2006). Other have proposed galaxy merger induced starbursts and subsequent supermassive blackhole growth as possible mechanisms (Springel et al. 2005; Di Matteo et al. 2005; Hopkins & Beacom 2006b; Hopkins et al. 2008a,b). Feedback from accreting active galactic nuclei (AGN) has also been suggested to contribute to quenching (sometimes in conjunction with other mechanisms; Croton et al. 2006; Cattaneo et al. 2006; Gabor et al. 2011); so has internal morphological instabilities in the galactic disk or bar (Cole et al. 2000; Martig et al. 2009). With so many proposed mechanisms available, observational constraints are critical to test them.

Several works have utilized the observed global trends of galaxy populations in order to construct empirical models for galaxy star formation histories and quenching (e.g. Wetzel et al. 2013; Schawinski et al. 2014; Smethurst et al. 2015). Central galaxies constitute over 70% of the $\mathcal{M}_* > 10^{9.7}\mathrm{M}_\odot$ galaxy population at $z = 0$. Moreover, the majority of the quiescent population at $\mathcal{M}_* > 10^{10}\mathrm{M}_\odot$ become quiescent as centrals (Wetzel et al. 2013). The quenching of central galaxies plays a critical role in the evolution of massive galaxies. In this paper, we take a similar approach as Wetzel et al. (2013) but for central galaxies. Wetzel et al. (2013) quantify the star formation histories and quenching timescales in a statistical and empirical manner. Then using the observed SSFR distribution of satellite galaxies, they constrain the quenching timescale of satellites and illustrate the success of a "delay-then-rapid" quenching model, where a satellite begins to quench rapidly only after a significant delay time after it infalls onto its central halo.

Extending to centrals, we use the global trends of the central galaxy population at $z < 1$ in order to construct a similarly statistical and empirical model for the star formation histories of central galaxies. While the initial conditions of the satellite galaxies in Wetzel et al. (2013) (at the times of their infall) are taken from observed trends of the central galaxy population, our model for central galaxies must actually reproduce all of the multifaceted observations. This requires us to construct a more comprehensive model that marries all the significant observational trends. Then by comparing the mock catalogs generated using our model to observations, we constrain the star formation histories and quenching timescales of central galaxies. Quantifying the timescales of the physical mechanisms that quench star formation, not only gives us a means for discerning the numerous different proposed mechanisms, but it also provides important insights into the overall evolution of galaxies.

We begin first in §4.3 by describing the observed central galaxy catalog at $z \approx 0$ that we construct from SDSS Data Release 7. Next, we describe the cosmological N-body simulation used to create a central galaxy mock catalog in §4.4. We then develop parameterizations of the observed global trends of the galaxy population and describe how we incorporate them into the mock catalog in §4.5. In §4.6, we describe how we use our model and the observed central galaxy catalog in order to infer the quenching timescale of central galaxies. Finally in §4.7 and §4.8 we discuss the implications of our results and summarize them.

4.3 Central Galaxies of SDSS DR7

We start by selecting a volume-limited sample of galaxies with $M_r - 5\log(h) < -18$ from the NYU Value-Added Galaxy Catalog (VAGC; Blanton et al. 2005b) of the Sloan Digital Sky Survey Data Release 7 (Abazajian et al. 2009) at redshift $z \approx 0.04$ following the sample selection of Tinker et al. (2011). The galaxy stellar masses are estimated using the `kcorrect` code (Blanton & Roweis 2007) assuming a Chabrier (2003) initial mass function (IMF). For measurement of galaxy star-formation, we use the specific star formation rate (SSFR) from the current release of Brinchmann et al. (2004) [1]. Generally, SSFRs $\gtrsim 10^{-11}\text{yr}^{-1}$ are derived

[1] http://www.mpa-garching.mpg.de/SDSS/DR7/

from Hα emissions, $10^{-11} \gtrsim$ SSFRs $\gtrsim 10^{-12}\text{yr}^{-1}$ are derived from a combination of emission lines, and SSFRs $\lesssim 10^{-12}\text{yr}^{-1}$ are mainly based on D_n4000 (see discussion in Wetzel et al. 2013). The spectroscopically derived SSFRs, which accounts for dust-reddening, allow us to make more accurate distinctions between star-forming and quiescent galaxies than simple color cuts. We note that SSFRs $\lesssim 10^{-12}\text{yr}^{-1}$ should only be considered upper limits to the true value (Salim et al. 2007).

Next, we identify the central galaxies using the halo-based group-finding algorithm from Tinker et al. (2011). For a detailed description we refer readers to Tinker et al. (2011); Wetzel et al. (2012, 2013, 2014), and Tinker et al. (2016). The most massive galaxy of the group is the 'central' galaxy and the rest are 'satellite' galaxies. In any group finding algorithm there are misassignments due to projection effects and redshift space distortions. Campbell et al. (2015), quantify both the purity and completeness of centrals identified using this group-finding algorithm at $\sim 80\%$. More importantly, they find that the algorithm can robustly identify red and blue centrals and satellites as a function of stellar mass and yield a nearly unbiased central red fraction, which is the key statistic relevant to our analysis here.

4.4 Simulated Central Galaxy Catalog

If we are to understand how central galaxies and their star formation evolve, we require simulations over a wide redshift range that allows us to examine and track central galaxies within the heirarchical growth of their host halos. To do this robustly, we require a cosmological N-body simulation that accounts for the complex dynamical processes that govern galaxy host halos. In this paper, we use the dissipationless, N-body simulation from Wetzel et al. (2013) generated using the White (2002) `TreePM` code with flat, ΛCDM cosmology: $\Omega_\text{m} = 0.274$, $\Omega_\text{b} = 0.0457$, $h = 0.7$, $n = 0.95$, and $\sigma_8 = 0.8$. 2048^3 particles are evolved in a 250 Mpc/h box with particle mass of $1.98 \times 10^8 \mathrm{M}_\odot$ and with a Plummer equivalent smoothing of 2.5 kpc/h. The initial conditions of the simulation at $z = 150$ are generated using second-order Lagrangian Perturbation Theory. We refer readers to Wetzel et al. (2013) and Wetzel et al. (2014) for a more detailed description of the simulation.

From the `TreePM` simulation, Wetzel et al. (2013) identify 'host halos' using the Friends-of-Friends (FoF) algorithm of Davis et al. (1985b) with linking length $b = 0.168$ times the mean inter-particle spacing. This groups the simulation particles bound by an isodensity contour of $\sim 100\times$ the mean matter density. Within the identified host halos, the simulation identifies 'subhalos' as overdensities in phase space through a 6-dimensional FoF algorithm (White et al. 2010). Wetzel et al. (2013) then track the host halos and subhalos across the simulation outputs in order to build merger trees. Next, Wetzel et al. (2013) designate the most massive subhalo in a newly-formed host halo at a given simulation out as the 'central' subhalo. A subhalo remains central until it falls into a more massive host halo, at which point it becomes a 'satellite' subhalo. Each subhalo is also assigned a maximum mass M_peak, the maximum host halo mass the subhalo has had in its history.

Using the Wetzel et al. (2013) simulation, we obtain a galaxy catalog from the subhalo catalog by assuming that galaxies reside at the centers of the subhalos and through subhalo abundance matching (SHAM; Vale & Ostriker 2006; Conroy et al. 2006; Yang et al. 2009; Wetzel et al. 2012; Leja et al. 2013; Wetzel et al. 2013, 2014) to assign them stellar masses. SHAM assumes a one-to-one mapping that preserves the rank ordering between subhalo M_peak and stellar mass, $\mathcal{M}_*$ of its galaxy: $n(> M_\mathrm{peak}) = n(> \mathcal{M}_*)$. Through SHAM, we can assign galaxy stellar masses to subhalos based on observed stellar mass function (SMF) at the redshifts of the simulation outputs. Galaxy stellar masses are assigned independently at each snapshot. This allows us to not only track the history of the subhalo, but also track the evolution of galaxy stellar masses through their SHAM stellar masses at each snapshot.

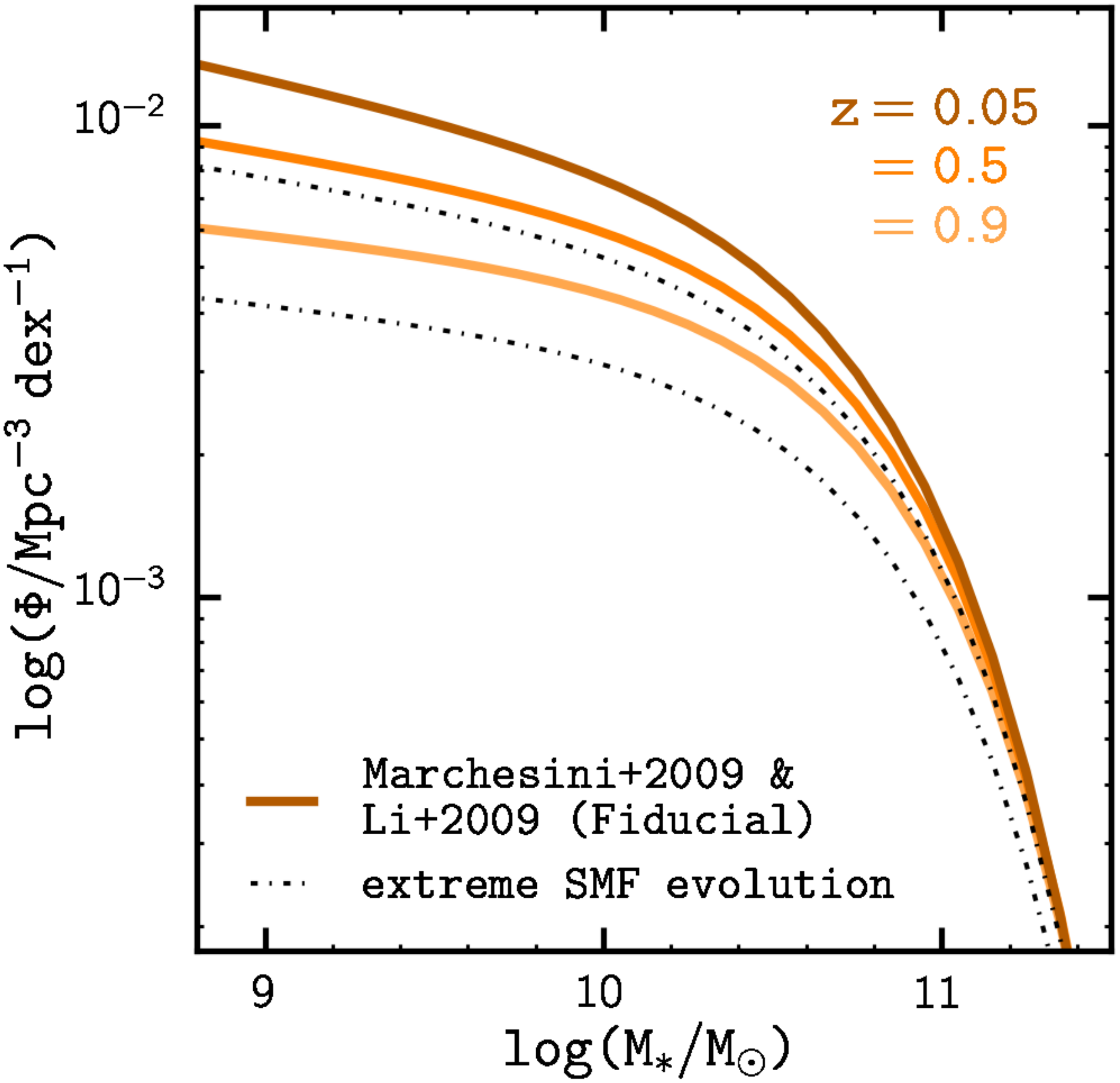

Figure 4.1: The stellar mass function (SMF) that we use in our subhalo abundance matching (SHAM) prescription to construct galaxy catalogs from the Wetzel et al. (2013) `TreePM` simulation (§4.4). For our fiducial SMF (solid), we use the Li & White (2009) SMF at $z = 0.05$ and interpolate between the Li & White (2009) SMF and the Marchesini et al. (2009) $z = 1.6$ SMF for $z > 0.05$. To illustrate the evolution, we plot the SMF at $z = 0.05, 0.5,$ and 0.9. We also plot a SMF parameterization using an "extreme" model of SMF evolution (dashed-dotted), in which the amplitude of the SMF at $z = 1.2$ is half the amplitude of the fiducial SMF. We later use this extreme model to ensure that the results in this work remain robust over different degrees of SMF evolution at $z > 0.05$.

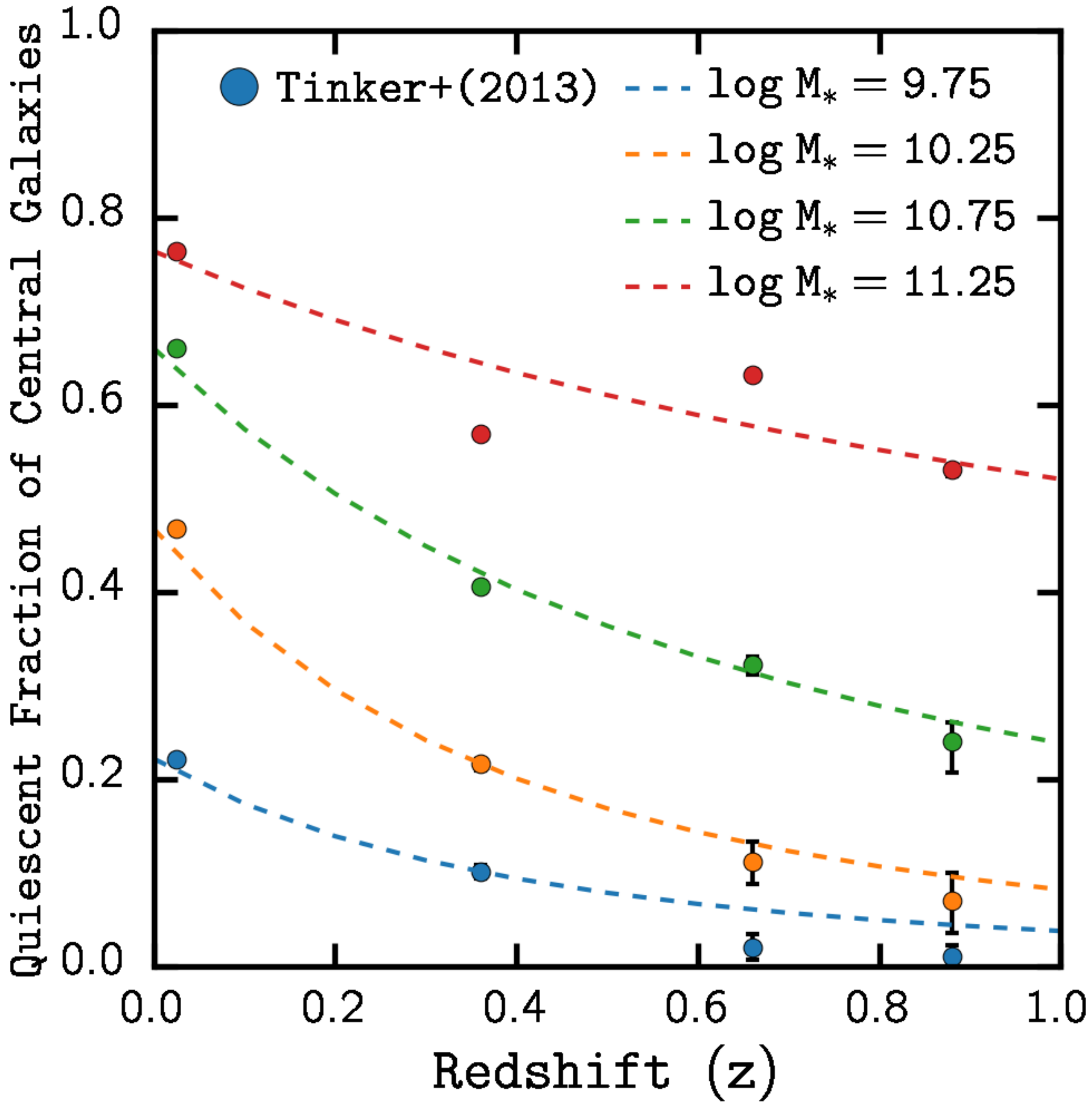

Figure 4.2: The quiescent fraction of central galaxies, $f_{\rm Q}^{\rm cen}$, at $z < 1$ in different stellar mass bins. We compare our parameterzation of $f_{\rm Q}^{\rm cen}$ (Eq. 4.2) using the best-fit parameter values listed in Table 4.1 (dashed) to the $f_{\rm Q}^{\rm cen}$ measurements from Tinker et al. (2013) (scatter). For our parameterization, we fit $f_{\rm Q}^{\rm cen}$ at $z = 0.$ using central galaxies of the SDSS DR7 group catalog and fit $\alpha(\mathcal{M}_*)$, which dictates the redshift dependence, from the redshift evolution of the Tinker et al. (2013) measurements.

For our SHAM prescription, we use the SMF of Li & White (2009) at the lowest redshift $z = 0.05$. Li & White (2009) is based on the same SDSS NYU-VAGC sample as the SDSS DR7 group catalog we describe in §4.3. At higher redshifts, we interpolate between the Li & White (2009) SMF and the Marchesini et al. (2009) SMF at $z = 1.6$ to obtain the SMF at the simulation output redshifts. This produces SMFs that increase significantly and monotonically over $z < 1$ for $\mathcal{M}_* < 10^{11}\mathrm{M}_\odot$ but insignificantly for $\mathcal{M}_* > 10^{11}\mathrm{M}_\odot$. We choose the Marchesini et al. (2009) SMF, amongst others, because it produces interpolated SMFs that monotonically increase at $z < 1$. At $z \sim 1$, the interpolated SMF we use is consistent (within the 1σ uncertainties) with more recent measurements from Muzzin et al. (2013) and Ilbert et al. (2013).

In Figure 4.1, we illustrate the evolution of the SMFs that we use for our SHAM prescription (solid) for $z = 0.05, 0.5$, and 0.9. Recently, using PRIMUS, Moustakas et al. (2013) found little evolution in the SMF for $z < 1$ at all mass ranges. Although previous works such as Bundy et al. (2006a) find otherwise. To ensure that our results do not depend on our choice of the SMFs, later in §4.6.2, we repeat our analysis using SMFs with no evolution (i.e. Li & White 2009 SMF throughout $0 < z < 1$) and with "extreme" evolution for $z > 0.05$ (dash-dotted in Figure 4.1), in which the amplitude of the SMF at $z = 1.2$ is approximately half the amplitude of the fiducial SMF at $z = 1.2$. Furthermore, while the simplest version of SHAM assumes a one-to-one correspondence between M_peak and $\mathcal{M}_*$, observations suggest that there is a scatter of ~ 0.2 dex in this relation (Zheng et al. 2007b; Yang et al. 2008; More et al. 2009b; Gu et al. 2016). Hence, we apply a 0.2 dex log-normal scatter in $\mathcal{M}_*$ at fixed M_peak in our SHAM prescription at each snapshot independently.

So far, we have subhalos populated with galaxies and their stellar mass at each of the 15 simulation outputs spanning the redshift $0.05 < z < 1$. For our sample, we restrict ourselves to galaxies classified as centrals by the simulation. And also to ones that are in both the $z \sim 0.05$ and $z \sim 1$ snapshots. This removes $< 3\%$ of central galaxies with $\mathcal{M}_* > 10^{9.5}$ $\mathrm{M}_\odot$ in the $z \sim 0.05$ snapshot. Our sample inevitably includes "back splash" or "ejected" satellite galaxies (Wetzel et al., 2014), misclassified as centrals. Excluding these galaxies, however, has a negligible impact on our results. We also note that while we do not have an explicit prescription for stellar mass growth from mergers, based on SHAM, the stellar mass growth traces the merger induced subhalo growth. As we discuss later in detail, mounting evidence disfavor merger driven quenching as the trigger of star formation quenching, so our treatment of mergers do not impact our quenching timescale results. In summary, we construct from our simulation a catalog of central galaxies whose stellar mass and halo mass is traced through the redshift range $0.05 < z < 1$.

4.5 Star Formation in Central Galaxies

The `TreePM` simulation (§4.4) provides a framework to examine the evolution of central galaxies within the ΛCDM hierarchical structure formation of the Universe. In order to determine the quenching timescale of central galaxies, we incorporate the evolution of star-formation within this framework so that the star formation of the simulated central galaxies reproduce observed trends. More specifically, we implement star formation in central galaxies to reproduce the observed evolution of the quiescent fraction and star-forming main sequence.

We begin in §4.5.1 by describing our paramertization of the observed quiescent fraction and SFMS evolutionary trends at $z < 1$. Afterwards, we describe the initial SFR assignment of the central galaxies in the simulation at the $z = 1$ snapshot in §4.5.2. Then in §4.5.3 we describe how our model evolves the SFRs and quenches these central galaxies.

4.5.1 Observations

With galaxy surveys like the SDSS, COSMOS, and PRIMUS, observations have firmly established that for $z < 2$, galactic properties such as color and star formation rate (SFR) have a bimodal distribution (Baldry et al. 2006; Cooper et al. 2007; Blanton & Moustakas 2009a; Moustakas et al. 2013). As mentioned above, the two main components of this distribution are *quiescent* galaxies with little star formation, which are redder, more massive, and reside in denser environments and *star forming* galaxies, which are bluer, less massive and more often found in the field. Since this bimodality is most likely a result of star formation being quenched in galaxies, measurements of the quiescent fraction f_{Q}, the fraction of quiescent galaxies in a population, is often used to indicate the overall star-forming property of galaxy populations (Baldry et al. 2006; Drory et al. 2009; Cooper et al. 2010; Iovino et al. 2010; Peng et al. 2010; Geha et al. 2012b; Kovač et al. 2014; Hahn et al. 2015).

For $z < 1$, observations find that the overall quiescent fraction increases as a function of stellar mass and with lower redshift (Drory et al. 2009; Iovino et al. 2010; Peng et al. 2010; Kovač et al. 2014; Hahn et al. 2015). In Wetzel et al. (2013), they quantify this mass and redshift dependence of the quiescent fraction through the parameterization, $f_{\mathrm{Q}}(\mathcal{M}_*, z) = A(\mathcal{M}_*) \times (1+z)^{\alpha(\mathcal{M}_*)}$, with $A(\mathcal{M}_*)$ and $\alpha(\mathcal{M}_*)$ fit from the quiescent fractions of the SDSS DR 7 catalog and the COSMOS survey at $z < 1$ (Drory et al. 2009), respectively. However, the quiescent fraction evolution is not universal over all environments (Hahn et al. 2015). More specifically, Tinker & Wetzel (2010) and Tinker et al. (2013) find distinct quiescent fraction evolutions for central and satellite galaxies.

We focus solely on the central galaxy quiescent fraction. We use the same parameterization as the overall quiescent fraction parameterization in Wetzel et al. (2013):

$$f_{\mathrm{Q}}^{\mathrm{cen}}(\mathcal{M}_*, z) = f_{\mathrm{Q}}^{\mathrm{cen}}(\mathcal{M}_*, z=0) \times (1+z)^{\alpha(\mathcal{M}_*)} \tag{4.1}$$

where

$$f_{\rm Q}^{\rm cen}(\mathcal{M}_*, z=0.) = A_0 + A_1\ \log \mathcal{M}_*, \tag{4.2}$$

is fit to the $f_{\rm Q}^{\rm cen}$ measured from SDSS DR7 group catalog central galaxies (§4.3) using a $\mathrm{SFR} - \mathcal{M}_*$ classification later described in § 4.6.1 (Eq. 4.18). $\alpha(\mathcal{M}_*)$, which dictates the redshift dependence of Eq. 4.1, is fit using the redshift dependence of $f_{\rm Q}^{\rm cen}$ measurements from Tinker et al. (2013), derived from observations of the SMF, galaxy clustering, and galaxy-galaxy lensing within the COSMOS survey, in bins of width $\Delta \log \mathcal{M}_* = 0.5$ dex. In Table 4.1, we list the best fit values for the parameters in Eq. 4.1 and in Figure 4.2 we compare our parameterization to the Tinker et al. (2013) measurements.

We note that Tinker et al. (2013) classifies galaxies as star-forming or quiescent based on a $(NUV - R) - (R - J)$ color-color cut from Bundy et al. (2010) rather than a $\mathrm{SFR} - \mathcal{M}_*$ classification such as Eq. 4.18. In Appendix 4.8 we confirm that for the SDSS DR7 group catalog, $f_{\rm Q}$ calculated using an $(NUV - R) - (R - J)$ color-color classification is consistent with $f_{\rm Q}$ calculated using a $\mathrm{SFR} - \mathcal{M}_*$ classification.

Observations of galaxy populations also find a tight correlation between the SFRs of star-forming galaxies and their stellar masses, which is referred to in the literature as the "star formation main sequence" (SFMS; Noeske et al. 2007; Oliver et al. 2010; Karim et al. 2011; Moustakas et al. 2013). Star-forming galaxies with higher stellar masses have higher SFRs. Roughly, this mass dependence can be characterized by a power law, $\mathrm{SFR} \propto \mathcal{M}^{\beta}$ and for a given stellar mass, SFRs follows a log-normal distribution (Noeske et al. 2007; Lee et al. 2015). Over cosmic time, this tight correlation decreases in SFR but has a constant scatter with $\sigma_{\log \mathrm{SFR}} \sim 0.3$ dex (Noeske et al. 2007; Elbaz et al. 2007; Daddi et al. 2007; Salim et al. 2007; Whitaker et al. 2012; Lee et al. 2015), In fact this decline SFR of star-forming galaxies in the SFMS is likely responsible for the remarkable decline of star formation in the Universe (Hopkins & Beacom 2006b; Behroozi et al. 2013b; Madau & Dickinson 2014).

Following the typical power-law parameterization of the SFMS, we construct a flexible parameterization that depends on mass and redshift. For a given stellar mass and redshift, the mean SFR of the SFMS is given by

$$\overline{\mathrm{SFR}}_{\mathrm{MS}}(\mathcal{M}_*, z) = A_{\mathrm{SDSS}} \left(\frac{\mathcal{M}_*}{10^{10.5}\mathrm{M}_\odot}\right)^{\beta_\mathcal{M}} 10^{\beta_z (z-0.05)}. \tag{4.3}$$

A_{SDSS} is the SFR of the SFMS for the SDSS group catalog at $\mathcal{M}_* = 10^{10.5}\mathrm{M}_\odot$. We determine $\beta_\mathcal{M}$ from fitting Eq. 4.3 to the SFMS of the SDSS group catalog ($z = 0.05$). Then we determine β_z such that the redshift dependence of our estimate of cosmic star formation rate,

$$\rho_{\mathrm{SFR}}(z) \propto \int \left(1 - f_{\mathrm{Q}}^{\mathrm{cen}}\right) \mathrm{SFR}_{\mathrm{MS}}(\mathcal{M}, z)\Phi(\mathcal{M}, z)\, \mathrm{d}\mathcal{M}, \tag{4.4}$$

is consistent with the redshift dependence of the cosmic star formation rate observations at $z < 1$ (Behroozi et al. 2013b). $\Phi(\mathcal{M}, z)$ and $f_{\mathrm{Q}}^{\mathrm{cen}}$ in Eq. 4.4, are the SMF used in the SHAM procedure (§4.4) and the central galaxy quiescent fraction (Eq. 4.1). This agreement in redshift dependence ensures the observational consistency between the SMF and the cosmic star formation density evolution, which Behroozi et al. (2013b) find. We list the best fit values to $A_{\mathrm{SDSS}}, \beta_\mathcal{M}$, and β_z in Table 4.1. Our $\beta_\mathcal{M}$ and β_z values are consistent with similar parameterizations in the literature (Salim et al. 2007; Moustakas et al. 2013; Lee et al. 2015).

4.5.2 Assigning Star Formation Rates

The first output of the `TreePM` simulation that we utilize is at $z_{\mathrm{initial}} = 1.08$. We designate the central galaxies of this snapshot as quenching, star-forming or quiescent and assign SFRs to them as the initial conditions of our model. The SFR assignment are based on the observed galaxy bimodality, the SFMS, and quiescent fraction at z_{initial} as we detail below.

First, we classify a fraction of the central galaxies as "*quenching*" galaxies – galaxies

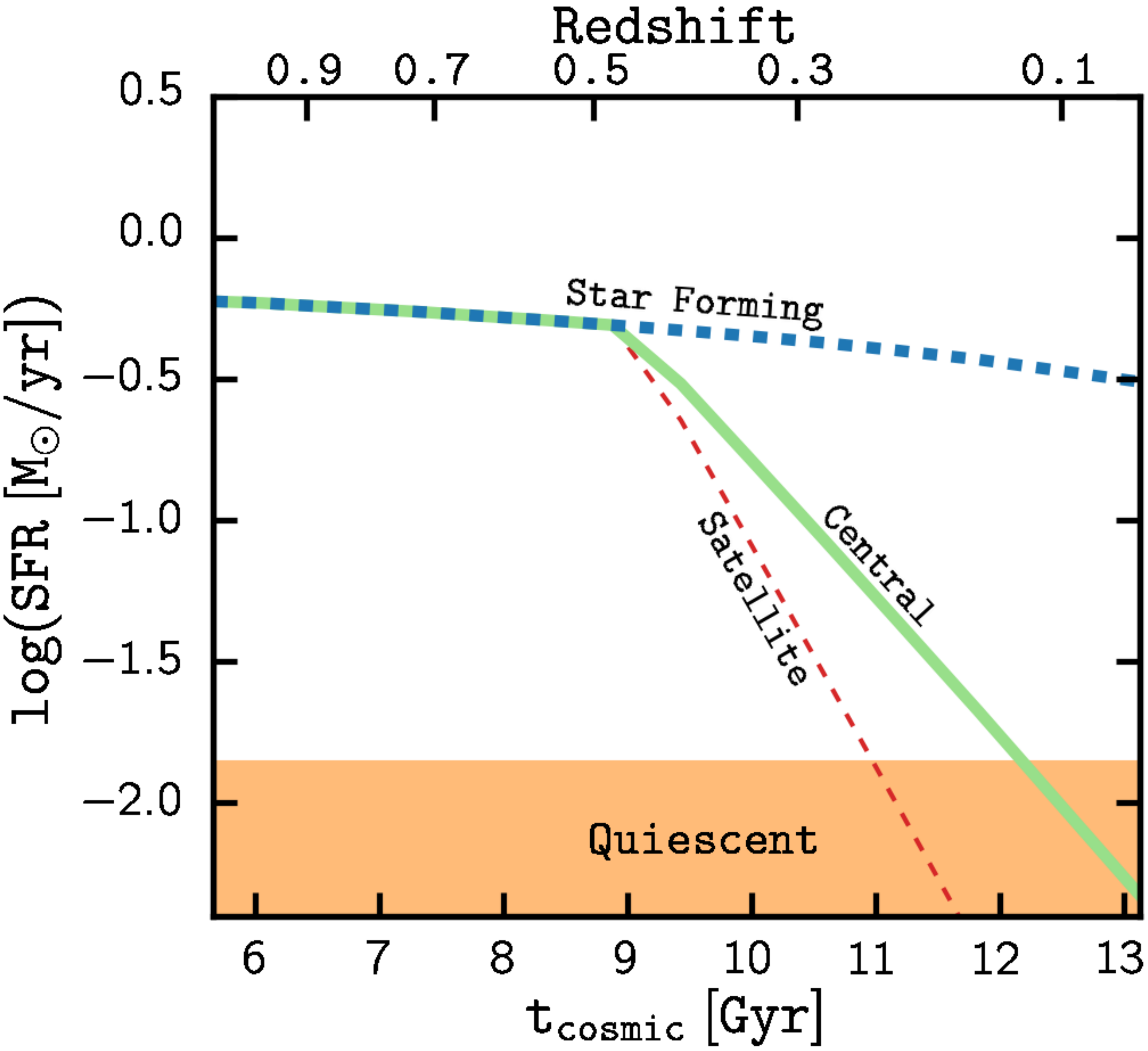

Figure 4.3: Schematic diagram that illustrate the star formation evolution of central galaxies in our model (§4.5.3). We plot SFR as a function of $t_{\rm cosmic}$ for a star forming galaxy (blue dashed), a star-forming galaxy that quenches at $t_{\rm Q,start} = 9$ Gyr (green) and mark the general region of quiescent galaxies (orange). Central galaxies while they are star-forming have SFRs that evolve with the SFMS, which decreases with cosmic time. When star-forming central galaxies quench, their SFR decreases exponentially with $t_{\rm cosmic}$. The quenching timescale, $\tau_{\rm Q}^{\rm cen}$, we constrain in our analysis dictates how rapidly these galaxies quench based on Eq. 4.8. For comparison we also plot the SFR evolution of a satellite with the same mass using the Wetzel et al. (2013) quenching timescales (red dashed).

that reside in the green valley, which are in the transitional state of becoming quiescent from star-forming. Current observations do not provide strong constraints on the fraction of galaxies "quenching" at $z \sim 1$. So, we use a flexible and mass dependent prescription

$$f_{GV}(\mathcal{M}_*) = A_{GV}(\log\ \mathcal{M}_* - 10.5) + \delta_{GV} \tag{4.5}$$

and marginalize over the nuisance parameters, A_{GV} and δ_{GV}, in our analysis. For these designated quenching central galaxies, we assign SFRs by uniformly sampling between the average SFR of the SFMS (Eq. 4.3) at $z_{\rm initial}$ and $\mathrm{SFR_Q}(\mathcal{M}_*)$, the SFR of the quiescent peak of the SDSS central galaxy SSFR distribution, which we later detail in this section.

Next, we classify the remaining $1 - f_{\rm GV}$ of the galaxy population as either star-forming or quiescent to match $f_{\rm Q}^{\rm cen}(z = z_{\rm initial})$. Galaxies classified as star-forming, are assigned SFRs based on the log normal SFR distribution of the SFMS at $z_{\rm initial}$ with scatter $\sigma_{\log \mathrm{SFR}} \sim 0.3$ (§4.5.1):

$$\log \mathrm{SFR}_{\rm SF}^{init} = \mathcal{N}(\log \overline{\mathrm{SFR}}_{\rm MS}\,(\mathcal{M}_*, z_{\rm initial}),\ 0.3)\,. \tag{4.6}$$

where $\mathcal{N}$ represents a Gaussian. Galaxies classified as quiescent are assigned SFRs based on a log-normal distribution centered about $\overline{\mathrm{SFR}}_{\mathrm{Q},init}$ with scatter $\sigma^{\rm Q}_{\log \mathrm{SFR}}$:

$$\log \mathrm{SFR}_{\rm Q}^{init} = \mathcal{N}(\overline{\mathrm{SFR}}_{\mathrm{Q},init}, \sigma^{\rm Q}_{\log \mathrm{SFR}}) \tag{4.7}$$

Both $\overline{\mathrm{SFR}}_{\mathrm{Q},init}$ and $\sigma^{\rm Q}_{\log \mathrm{SFR}}$ are determined empirically from the quiescent peak SSFR in the SDSS central galaxy SSFR distribution: $\overline{\mathrm{SFR}}_{\mathrm{Q},init} = 0.4\ (\log \mathcal{M}_* - 10.5) - 1.73$ and $\sigma^{\rm Q}_{\log \mathrm{SFR}} = 0.18$. Our aim is solely to empirically reproduce the quiescent peak because the SSFR measurements are largely upper limits, so the peak itself is nonphysical (§4.3).

4.5.3 Star Formation Evolution

Starting from the initial SFRs of the central galaxies that we just assigned, next, we evolve the SFRs in order to reproduce the observed evolution of the quiescent fraction and the SFMS (§4.5.1). The central galaxies in our simulation evolve their SFRs as star-forming galaxies, quiescent galaxies and quench their star-formation. With the focus of this work on the quenching timescale of central galaxies, we first discuss how we evolve the SFRs of central galaxies that quench within our simulation. Then we discuss how we evolve the SFRs of central galaxies while they are star-forming and after they have quenched their star formation.

Once a galaxy begins to quench its star formation, its SFR decreases and, on the SFR-$\mathcal{M}_*$ relation, it migrates from the SFMS to the quenched sequence. We designate the time when a galaxy starts to quench as $t_{Q,\mathrm{start}}$ and model its decline in SFR exponentially with a characteristic e-folding time τ_Q^{cen}, which we refer to as the "central quenching timescale":

$$\mathrm{SFR}_{\mathrm{Quenching}}(t) = \mathrm{SFR}_{\mathrm{SF}}(t) \times \exp\left(-\frac{t - t_{Q,\mathrm{start}}}{\tau_Q^{\mathrm{cen}}}\right). \tag{4.8}$$

$\mathrm{SFR}_{\mathrm{SF}}$ represents the SFR of a star-forming central galaxy, which we define later, and τ_Q^{cen} characterizes how long quenching mechanism(s) take(s) to cease star-formation in a central galaxy. In order to determine whether this timescale depends on the stellar mass of the galaxy, we include a mass dependence:

$$\tau_Q^{\mathrm{cen}}(\mathcal{M}_*) = A_\tau\,(\log\ \mathcal{M}_* - 11.1) + \delta_\tau. \tag{4.9}$$

In addition to the SFR evolution after they begin to quench, our model must also quantify when and how many star-forming centrals quench from z_{initial}.

For satellite galaxies, the moment they start quenching can be related to the moment when their host halo is accreted into the central galaxy's host halo via a time delay of several Gyrs (Wetzel et al. 2013). However, for central galaxies, the time when they start to quench is likely characterized by more complex and stochastic mechanisms such as gas depletion from strangulation (Peng et al. 2015), hot gas quenching (Gabor et al. 2010; Gabor & Davé 2012, 2015) or the onset of AGN activity. Fortunately, using the evolution of the quiescent fraction, we can statistically model the number of star-forming centrals that quenches throughout the simulation. We use a Monte Carlo prescription that utilizes a "quenching probability" (P_Q) to determine which star-forming centrals to quench and when to quench them. We define $P_Q(\mathcal{M}_*, t_i)$ to be the probability that a star-forming central of stellar mass $\mathcal{M}_*$ begins to quench some time between t_i and t_{i+1}.

In the fiducial case where quenching happens instantaneously and the time evolution of the stellar mass function is negligible, the quenching probability is given directly by the derivative of the quiescent fraction over time:

$$P_Q^{fid}(\mathcal{M}, t_i) = \frac{t_{i+1} - t_i}{1 - f_{\mathrm{Q}}(\mathcal{M}, t_i)} \frac{\mathrm{d}f_{\mathrm{Q}}}{\mathrm{d}t}. \tag{4.10}$$

However, to account for the SMF evolution, we introduce a correction to Eq. 4.10:

$$\Delta P_Q(\mathcal{M}, t_i) = \frac{N_{\mathrm{tot}}(\mathcal{M}, t_{i+1}) - N_{\mathrm{tot}}(\mathcal{M}, t_i)}{N_{SF}(\mathcal{M}, t_i)} f_Q(\mathcal{M}, t_{i+1}). \tag{4.11}$$

Furthermore, star-forming galaxies do not quench instantaneously. This implies that some galaxies can begin their quenching but still have high enough SFRs to be misclassified as star-forming causing a discrepancy between when star-forming galaxies start quenching to when they become classified as quiescent. This discrepancy depends on the SFRs of the quenching galaxies and the timescales of the quenching mechanism. Our ultimate goal is to characterize this timescale and its dependence on stellar mass, so any strong assumptions may bias our results. Therefore, we include a flexible mass dependent factor parameterized by A_{P_Q} and δ_{P_Q} to the quenching probability prescription:

$$f_{P_Q}(\mathcal{M}) = A_{P_Q}(\log \mathcal{M} - 10.5) + \delta_{P_Q}. \tag{4.12}$$

By including this term to the quenching probability, we treat A_{P_Q} and δ_{P_Q} as nuisance parameters, which mitigate any biases. Combined, the quenching probability we use is

$$P_{Q,i}(\mathcal{M}) = f_{P_Q}\left(P^{fid}_{Q,i} + \Delta P_{Q,i}\right). \tag{4.13}$$

Later in §4.6.2 we discuss the potential impact of our quenching probability parameterization on our results. In practice, at each simulation output snapshot t_i, a number of star-forming central galaxies are selected to start quenching based on their assigned quenching probabilities. We note that our quenching probability prescription quenches star-forming galaxies anywhere on the SFMS.

For quenching galaxies before they begin the quenching process and for star-forming galaxies that remain star-forming throughout, their star formation histories are dictated by the evolution of the SFMS. Therefore, we model the star formation evolution of star-forming central galaxies to statistically trace the redshift and mass dependence of the SFMS. Recall that the stellar masses of the central galaxies evolve independently from their star formation histories. Through our SHAM prescription, the stellar mass growth traces the mass accretion of its host subhalo (§ 4.4). Then, for a star-forming central with initial stellar mass $\mathcal{M}_0$ at $z_{\rm initial}$ that evolves to $\mathcal{M}$ at z to remain on the SFMS, based on Eq. (4.3), the SFR at $z_{\rm initial}$

must evolve by the following factor

$$f_{\rm MS} = \left(\frac{\mathcal{M}}{\mathcal{M}_0}\right)^{\beta_\mathcal{M}} \times 10^{\beta_z (z - z_0)} \tag{4.14}$$

where $\beta_\mathcal{M}$ and β_z are the fixed parameters that characterize the mass and redshift dependence of the SFMS (§ 4.5.1 and Table 4.1). So while central galaxies with $\rm SFR_0$ at $z_{\rm initial}$ remain star-forming they have,

$$\rm SFR_{SF} = \mathit{f}_{MS} \times SFR_0. \tag{4.15}$$

This way, star-forming galaxies follow the observed redshift evolution and mass dependence of the SFMS. Furthermore, since our prescription keeps the relative positions in SFR from $\overline{\rm SFR}_{\rm MS}$ constant, it preserve the SFR scatter of the SFMS – matching observations.

Of course, in reality, the SFHs of star-forming central galaxies do not strictly follow a simple parameterization of the SFMS evolution. The stellar mass growth of the star-forming centrals is not only related to the growth of its host subhalo, as our SHAM prescription assumes, but also linked to their SFHs. Observations, however, suggest a non-trivial connection between stellar mass growth, SFH, and host subhalo growth. For instance, if we estimate the stellar masses of star forming galaxy by integrating SFRs over time, then the stellar mass growth of star-forming galaxies with the same initial stellar mass but different SFR on the SFMS, would diverge over time and the final stellar masses will be significantly different. In that case, it would be difficult to preserve the SFR scatter in SFMS along with its log-normal characteristic. Alternatively, if independent of subhalo growth, the SFH linked stellar mass growth would cause the stellar mass growth for fixed halo mass to diverge. This would violate the observed scatter in the Stellar Mass to Halo Mass (SMHM) relation (Leauthaud et al., 2012a; Tinker et al., 2013; Zu & Mandelbaum, 2015; Gu et al., 2016). Clearly a mechanism such as a "*star formation duty cycle*" is required to consolidate observations of the SMHM and the SFMS. For the scope of this paper, however, we find that our above prescription of statistically evolving the SFRs of star-forming galaxies is sufficient and incorporating stellar mass growth through integrated SFR with a stochastic star forming duty cycle, does not significant impact the constraints on the quenching timescales. We will investigate star formation duty cycle in star-forming central galaxies and the link between stellar mass growth, host halo growth and SFH in Hahn et al. in prep.

Lastly, central galaxies that are quiescent at $z_{\rm initial}$ or become quiescent during the simulation remain quiescent. Their SFR evolution is determined only to empirically reproduce the quiescent peak of the SSFR distribution at $z = 0.05$, similar to the initial SFR assignment in §4.5.2. For galaxies that are quiescent at $z_{\rm initial}$, we evolve the SFRs in order to conserve the SSFRs throughout the simulation: $\mathrm{SSFR_Q} = \mathrm{SSFR_0}$, the initial SSFR at $z_{\rm initial}$. Then,

$$\mathrm{SFR_Q} = \mathrm{SSFR_0} \times \mathcal{M}_* \tag{4.16}$$

where $\mathcal{M}_*$ is stellar mass at the simulation outputs derived from SHAM.

For galaxies that start off as star-forming at $z_{\rm initial}$ and quench during the simulation, based on Eq. 4.8 their SFRs can decrease enough so that their SSFRs fall below the SSFR upper bounds of the Brinchmann et al. (2004) SSFR measurements. When we later compare the SSFR distributions of our model to the SDSS DR7 central galaxy catalog, the quenching galaxies with SSFRs below the SSFR bounds will spuriously cause discrepancies in the comparison. To prevent this, we impose a final quenched SSFR, based the quiescent peak of the observed SSFR distribution, for each galaxy when it begins to quench. Therefore, in our model, SFR of quenching galaxies only decreases until an assigned final quenched SSFR. Afterwards, their SFRs are evolved to conserve the SSFR (Eq. 4.16).

Figure 4.3 qualitatively illustrates the SFR evolution of star-forming (blue dashed), quenching (green solid), and quiescent (orange) central galaxies as a function of cosmic time throughout the simulation. The SFR of star-forming galaxies reflects the SFR evolution of the SFMS which decreases with time. The quenching galaxy starts to quench at $t_{\mathrm{Q,start}} = 9$ Gyr. Its departure from the SFMS is clearly illustrated at $t_{\mathrm{cosmic}} > 9$ Gyr. The slope of its SFR decline is dictated by the quenching timescale. Since the lower bound of the SSFR in the SDSS group catalog does not have any physical significance, we broadly mark the region with $\log\ \mathrm{SSFR} < \log\ \mathrm{SSFR_Q} + \sigma^{\mathrm{Q}}_{\log\ \mathrm{SFR}}$ as quiescent in Figure 4.3.

More quantitatively, in the top panel of Figure 4.4 we present the evolution of the SSFR distribution in our model (for a reasonable set of model parameter values) to illustrate how we track the star formation of central galaxies from $z \sim 1$. For this particular set of model parameter values, $f_{GV} \sim 0$ within the stellar mass bin. We plot the SSFR distribution for central galaxies in the stellar mass range $[10^{10.1}\mathrm{M_\odot}, 10^{10.5}\mathrm{M_\odot}]$ for a number of simulation output snapshots in the redshift range $0 < z < 1$ (top; darker with time). It demonstrates how our model reproduces the observed evolution of the SFMS and quiescent fraction. With time, the star-forming peak of the SSFR distribution decreases in SSFR tracing the SFMS evolution. The amplitude of the star-forming peak also decreases and is accompanied by the growth of the quiescent peak, reflecting the quiescent fraction evolution and the lower bound of SSFR measurements we impose.

In the bottom panel, we compare the SSFR distribution of our model at $z = 0.05$ using a relatively shorter (dashed) and longer (dotted) quenching timescale than in the top panel (solid). The quenching timescale (parameterized by A_τ and δ_τ in Eq. 4.9) dictates how long quenching central galaxies take to migrate from the SFMS to quiescence. The comparison illustrates that the length of the quenching timescale is reflected in the "height" of the SSFR distribution green valley. Longer quenching timescales, result in a higher green valley. Shorter quenching timescales, result in a lower one.

Table 4.1: Parameterizations in the Central Galaxy SFH Model with Fixed Parameters

Parameter	Value
Central Galaxy Quiescent Fraction (Eq. 4.1)	
$f_{\mathrm{Q}}^{\mathrm{cen}}(\mathcal{M}_*, z) = f_{\mathrm{Q}}^{\mathrm{cen}}(\mathcal{M}_*, z=0) \times (1+z)^{\alpha(\mathcal{M}_*)}$ $= (A_0 + A_1 \log \mathcal{M}_*) \times (1+z)^{\alpha(\mathcal{M}_*)}$	
A_0	-6.04
A_1	0.64
$\alpha(\mathcal{M}_*)$	-2.57 $\mathcal{M}_* \in [10^{9.5} - 10^{10}\mathrm{M}_\odot]$
	-2.52 $\mathcal{M}_* \in [10^{10} - 10^{10.5}\mathrm{M}_\odot]$
	-1.47 $\mathcal{M}_* \in [10^{10.5} - 10^{11}\mathrm{M}_\odot]$
	-0.55 $\mathcal{M}_* \in [10^{11} - 10^{11.5}\mathrm{M}_\odot]$
	-0.12 $\mathcal{M}_* \in [10^{11.5} - 10^{12}\mathrm{M}_\odot]$
SFMS SFR z and $\mathcal{M}_*$ Dependence (Eq. 4.3)	
$\overline{\mathrm{SFR}}_{\mathrm{MS}} = A_{\mathrm{SDSS}} \left(\frac{\mathcal{M}_*}{10^{10.5}\mathrm{M}_\odot}\right)^{\beta_{\mathcal{M}}} 10^{\beta_z(z-0.05)}$	
A_{SDSS}	$10^{-0.11}$ $\mathrm{M}_\odot/\mathrm{yr}$
$\beta_{\mathcal{M}}$	0.53
β_z	1.1

We list the parameters and their best-fit values for the central galaxy quiescent fraction (Eq. 4.1) and SFMS SFR redshift and stellar mass dependence (Eq. 4.3) parameterizations. $f_{\mathrm{Q}}^{\mathrm{cen}}(z=0)$ is fit using the central galaxies of the SDSS DR7 group catalog and the redshift dependence is fit using $f_{\mathrm{Q}}^{\mathrm{cen}}$ measurements from Tinker et al. (2013). Similarly, $\overline{\mathrm{SFR}}_{\mathrm{MS}}(z = 0.05)$ is fit using the group catalog while the redshift dependence paramerization is fit to reproduce the redshift dependence of the Behroozi et al. (2013b) cosmic star formation at $z < 1$.

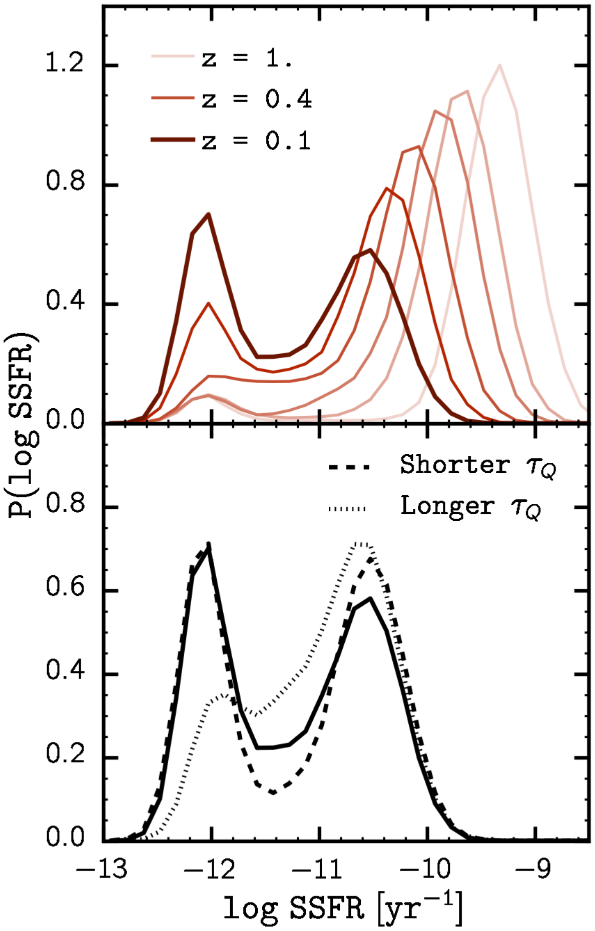

Figure 4.4: **Top**: The evolution of the SSFR distribution in our model (§4.5.2) for a reasonable set of parameter values. The model evolves the SFR of central galaxies from $z \sim 1$ (light) to 0.05 (dark) while reproducing the observed SFMS and quiescent fraction evolutions. The shift in the star forming peak of the SSFR distribution from $z = 1$, reflects the overall decline in SFR of the SFMS over time. The quiescent fraction evolution is reflected in the growth of the quiescent peak accompanied by the decline of the star-forming peak. **Bottom**: Comparison of the SSFR distribution at $z = 0.05$ using a relatively shorter (dashed) and longer (dotted) quenching timescale than the above panel (solid). The quenching timescale (parameterized by A_τ and δ_τ), dictates how long quenching central galaxies spend in between the peaks. This is ultimately reflected in the height of the green valley. For longer quenching timescales, the height of the SSFR distribution green valley will higher. For shorter quenching timescales, it will lower.

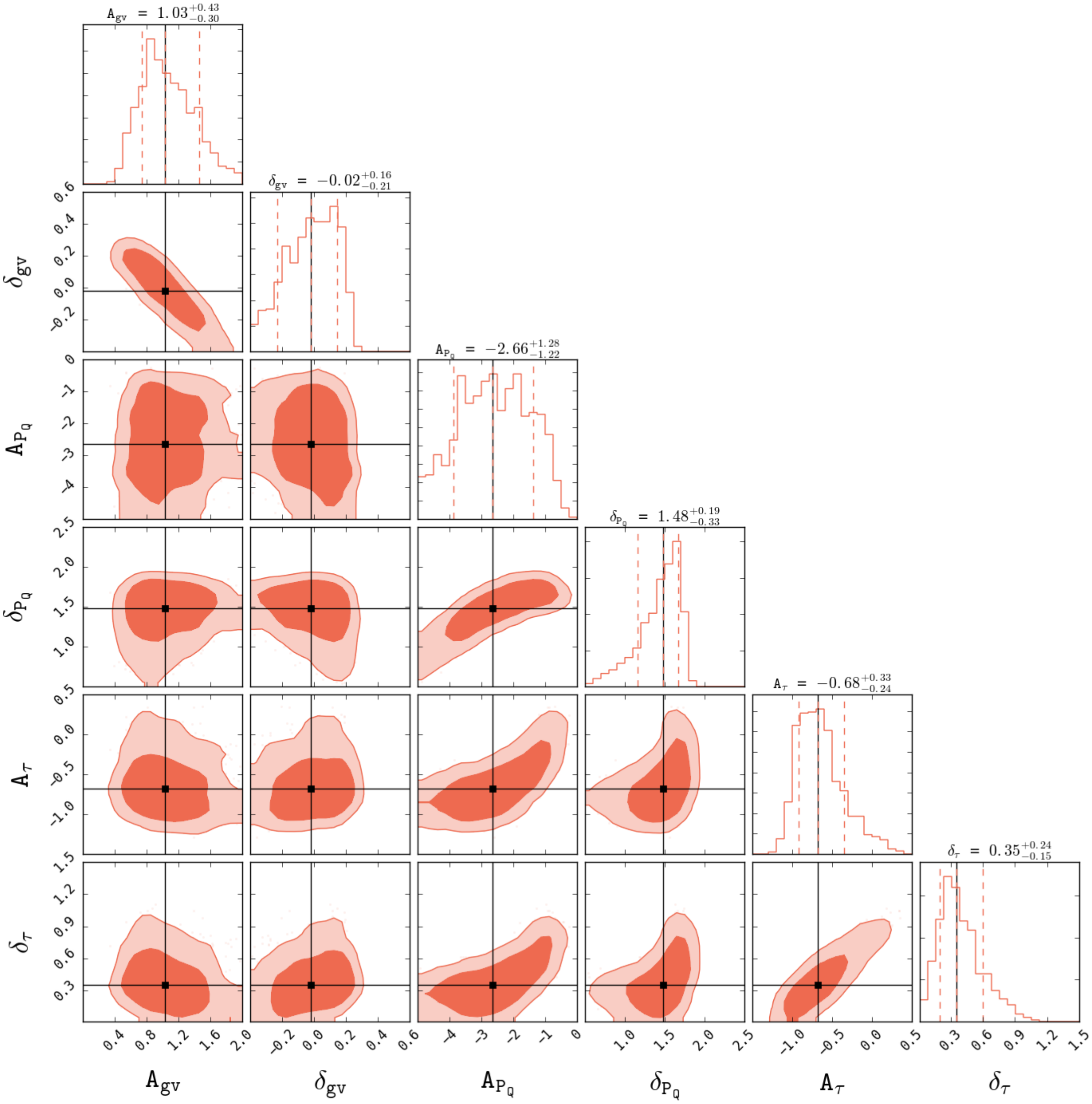

Figure 4.5: We present the constraints we obtain for our model parameters using ABC-PMC. The diagonal panels plot posterior distributions of each of our model parameters, while the off-diagonal panels plot the degeneracies of parameter pairs. For each of the posterior distributions, we mark the 68% confidence interval (vertical dashed lines). We also mark the median of the posterior distributions in all the panels in black.

Table 4.2: Parameterizations in the Central Galaxy SFH Model with Free Parameters

Quantity	Description	Parameter	Prior
$\tau_Q^{\rm cen} =$	Central Quenching Timescale	A_τ	$[-1.5, 0.5]$
$A_\tau(\log\ \mathcal{M}_* - 11.1) + \delta_\tau$	in Gyrs (Eq. 4.9)	δ_τ	$[0.01, 1.5]$
$f_{GV} =$	Initial $z \approx 1$ Green Valley	A_{GV}	$[0., 1.]$
$A_{GV}(\log\ \mathcal{M}_* - 10.5) + \delta_{GV}$	Fraction (Eq. 4.5)	δ_{GV}	$[-0.4, 0.6]$
$f_{P_Q} =$	Quenching Probability	A_{P_Q}	$[-5., 0.]$
$A_{P_Q}(\log \mathcal{M}_* - 10.5) + \delta_{P_Q}$	Factor (Eq. 4.12)	δ_τ	$[0.5, 2.5]$

We list the parameterizations of the quenching timescale (Eq. 4.9), the initial $z \approx 1$ green valley fraction (Eq. 4.5), and the quenching probability factor (Eq.4.12) that we use in our model (§4.5). In our Approximate Bayesian Computation parameter inference, we constrain the parameters listed in the four column. For the prior probability distributions of these parameters, we use uniform priors with the ranges listed in the last column. We note that while we allow $\delta_{GV} < 0$ due to the mass dependence of f_{GV}, f_{GV} can only be non-negative in our model.

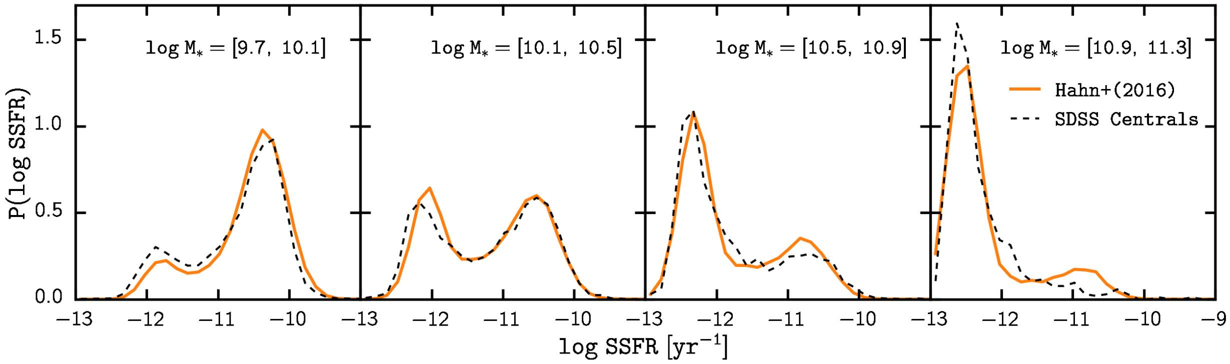

Figure 4.6: Comparison between the SSFR distribution calculated using the median of the ABC posterior distribution as the set of model parameters (orange) and the SSFR distribution of the SDSS DR7 central galaxies (black dash). The SSFR distribution from the median of the ABC posterior show good overall agreement. The distributions are especially consistent in the transition (green valley) regions, which are dictated by the quenching timescale parameters.

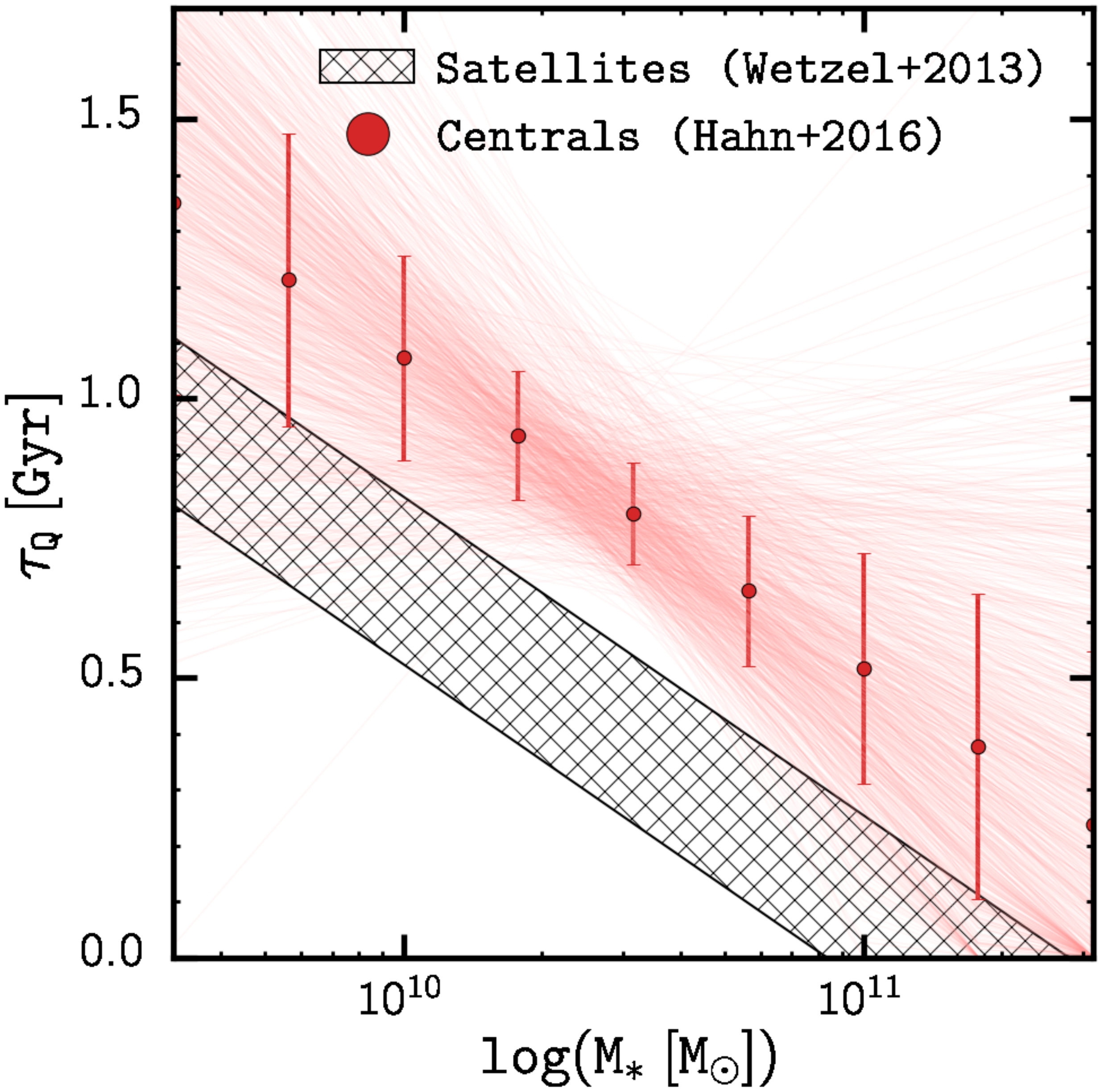

Figure 4.7: Quenching timescale, $\tau_Q^{\rm cen}$, of central galaxies (red) as a function of stellar mass. We plot $\tau_Q^{\rm cen}$ of the median parameter values of the ABC posterior distributions (red points) along with $\tau_Q^{\rm cen}$ drawn from the final iteration ABC parameter pool (faint red lines). For comparison, we also plot the satellite quenching timescale of Wetzel et al. (2014) (black dashed). The constraints we get for quenching timescale of central galaxies reveal that central galaxies have a significantly longer quenching timescale than satellite galaxies.

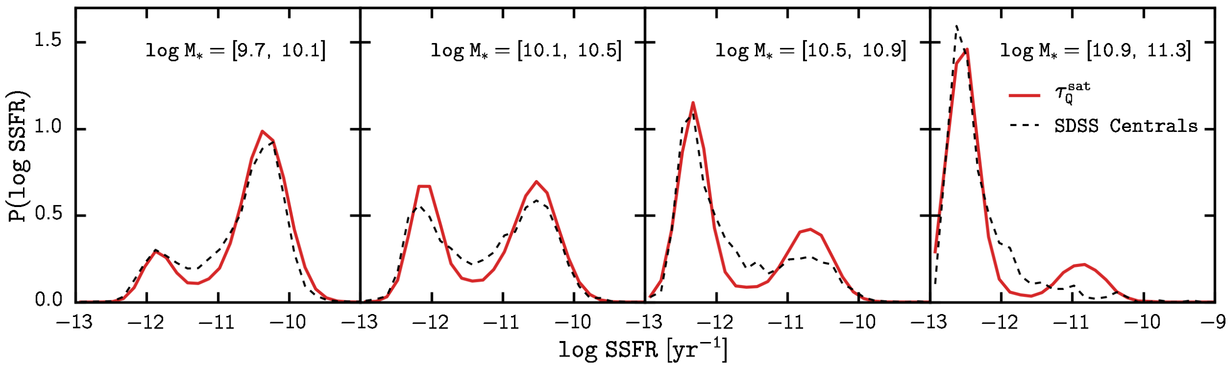

Figure 4.8: The SSFR distribution generated from the median values of the parameter constraints obtained from our analysis using Wetzel et al. (2013) satellite quenching timescale as the quenching timescale of our central galaxies (red) in four stellar mass bins. In each panel, we reproduce the quiescent fraction of the SDSS DR7 central galaxies; however, comparison to the SSFR distribution of the SDSS DR7 centrals (black dash) find significant discrepancies in each of the bins. The SSFR distribution using satellite quenching timescale have much shallower green valley regions as a result of galaxies quenching much faster with satellite quenching timescale. *This disagreement of model predictions for satellites applied to observations of centrals clearly demonstrates that centrals require longer quenching timescales than satellites.*

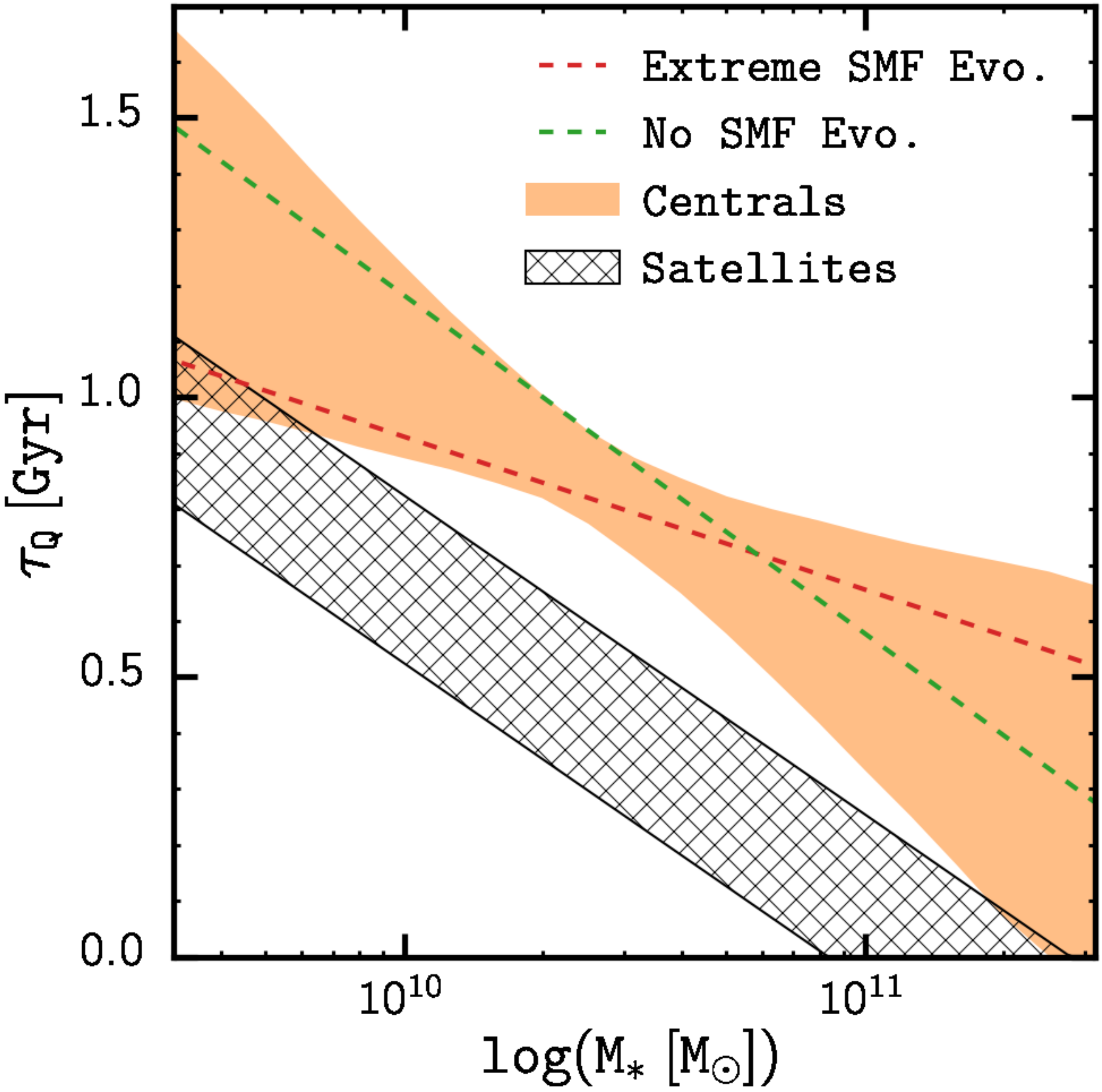

Figure 4.9: Central galaxy quenching timescales ($\tau_Q^{\rm cen}$) derived from using SMF prescriptions with no SMF evolution (green) and with extreme SMF evolution (red) in our analysis. For comparison we include the satellite galaxy quenching timescale from Wetzel et al. (2013) and $\tau_Q^{\rm cen}$ we obtain using our fiducial SMF prescription. Even extreme choices for the SMF evolution is insufficient to account for the significant difference between the central and satellite quenching timescales. The different SMF evolution mainly impacts the mass dependence, not the amplitude of $\tau_Q^{\rm cen}$.

4.6 Results

Now that we have a model for evolving star formation in central galaxies, in this section, we constrain the parameters of the model.

4.6.1 Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is a generative, simulation-based inference for robust parameter estimation. It has the advantage over standard approaches for parameter inference in that it does not require explicit knowledge of the likelihood function. It only relies on a simulation of observed data and on a metric for the distance between the observed data and simulation. It has already been effectively used for astronomy and cosmology in the literature (Cameron & Pettitt 2012b; Weyant et al. 2013b; Akeret et al. 2015b; Ishida et al. 2015b; Lin & Kilbinger 2015b; Lin et al. 2016b; Hahn et al. 2016, and Cisewski et al. in prep.), spanning a wide range of topics. For our purposes, which is to constrain the quenching timescale parameters, we use the observed SSFR distribution and quiescent fraction. ABC provides an ideal framework for parameter inference without having to specify the explicit likelihood of these observables. In practice, we use ABC in conjunction with the efficient Population Monte Carlo (PMC) importance sampling (Ishida et al. 2015b; Hahn et al. 2016).

ABC requires a number of specific choices for implementation: a simulation of the data, a set of prior probability distributions for the model parameters, and a distance metric to compare the "closeness" of the simulation to the data. In §4.5, we described our model for the star formation evolution of central galaxies. The parameters of our model, which we constrain in our ABC analysis are listed in Table 4.2. For the prior probability distributions of the simulation parameters, $\{A_{\mathrm{GV}}, \delta_{\mathrm{GV}}, A_{P_Q}, \delta_{P_Q}, A_{\tau}, \delta_{\tau}\}$, we choose uniform priors with conservative ranges also listed in Table 4.2.

The distance metric in ABC parameter estimation is — in principle — a positive definite function that compares various summary statistics between the data and the simulation. It can be a vector with multiple components where each component is a distance between one particular summary statistic of the data and that of the simulation. For our case, the summary statistics we use for our distance metric are the observables we seek to reproduce with our model: *the quiescent fraction evolution and SDSS DR7 central galaxy SSFR distribution.* Therefore, we use a two component distance metric, $\vec{\rho} = [\rho_{\rm QF}, \rho_{\rm SSFR}]$.

We calculate the first component, $\rho_{\rm QF}$, so that our model best reproduces the quiescent fraction at multiple snapshots:

$$\rho_{\rm QF} = \sum_{\mathcal{M}} \sum_{z' \in \{z\}} \big(f_{\rm Q}^{\rm cen}(\mathcal{M}, z') - f_{\rm Q}^{\rm model}(\mathcal{M}, z')\big)^2 \tag{4.17}$$

where $\{z\} = \{0.05, 0.16, 0.34, \text{ and } 1.08\}$ and $f_{\rm Q}^{\rm cen}$ is the parameterization of the observed quiescent fraction (Eq. 4.1). For $f_{\rm Q}^{\rm model}$ rather than using the evolutionary stages of the simulation central galaxies in the model, we measure it using the same $\mathrm{SFR} - \mathcal{M}_*$ classification used for deriving $f_{\rm Q}^{\rm cen}$ in Eq. 4.1 and 4.2. The $\mathrm{SFR} - \mathcal{M}_*$ cut in this classification is derived from the slope of the SFMS relation (Eq. 4.3):

$$\log\ \mathrm{SFR}_{\rm cut} = \log\ \overline{\mathrm{SFR}}_{\rm MS} - 0.9. \tag{4.18}$$

If a galaxy SFR is less than $\mathrm{SFR}_{\rm cut}$, then it is classified as quiescent; otherwise, as star-forming. This classification is analogous to the quiescent/star-forming classification of Moustakas et al. (2013), which also utilizes the slope of the SFMS. By measuring the quiescent fraction of the simulation we are more consistent with observations, which have no way of knowing the evolutionary stage of galaxies beyond their SFR and $\mathcal{M}_*$.

Our redshift choices for Eq. 4.17 is primarily motivated to ensure that our model agrees with the observed quiescent fraction throughout the lower redshifts ($z < 0.5$). By incorporating the $z' < 0.5$ contributions, we constrain $A_{\mathrm{P_Q}}$ and $\tau_{\mathrm{P_Q}}$, which dictate the quenching probabilities. $z_{\mathrm{initial}} = 1.08$ is also included to ensure that our initial conditions are consistent with observations.

The second component of our distance metric compares the SSFR distribution of the SDSS DR7 central galaxies to that of our model. More specifically,

$$\rho_{\mathrm{SSFR}} = \sum_{\mathrm{SSFR}} \left(P(\mathrm{SSFR})^{\mathrm{SDSS}} - P(\mathrm{SSFR})^{\mathrm{model}} \right)^2 . \tag{4.19}$$

As we discuss in § 4.5.3, the quenching timescale parameters leave an imprint on the SSFR distribution. So ρ_{SSFR} successfully serves to constrain A_τ and δ_τ.

We note that the low SSFR end of $P(\mathrm{SSFR})^{\mathrm{SDSS}}$ is impacted by the fact that the SSFR $\lesssim 10^{-12}\ \mathrm{yr}^{-1}$ from Brinchmann et al. (2004) are upper bounds (§ 4.3). ρ_{SSFR}, however, does not account for this effect. Instead, as we describe in § 4.5.3, we include this effect in our model when dealing with quiescent and quenching galaxies. Therefore, we do not expect the low SSFRs in Brinchmann et al. (2004) to significantly impact the constraints on A_τ and δ_τ.

Beyond our choice of distance metric, we strictly follow the ABC-PMC implementation of Hahn et al. (2016). For aficionados, we use a median distance threshold after each iteration of the PMC and declare convergence when the acceptance ratio falls below 1%. Once converged, the ABC algorithm produces parameter distributions that generate models with quiescent fractions and SSFR distributions close to observations. Moreover, these parameter distributions predict the posterior distributions of the parameters. For further details, we refer readers to Hahn et al. (2016).

4.6.2 Central Galaxy Quenching Timescale

We present the central galaxy quenching timescale constraints we obtain using ABC (§4.6.1), in Figure 4.5. The diagonal panels of the figure plot the posterior distribution of each of our model parameters with vertical dashed lines marking the median and the 68% confidence interval. The off-diagonal panels plot the degeneracies between parameter pairs. We also mark the median of the posterior distribution for each of the parameters (black). The off-diagonal panels illustrate that the initial green valley parameters are not degenerate with the other parameters. Galaxies that are initially in the green valley quickly evolve out of it, so the green valley prescription is mainly constrained by the quiescent fraction at $z_{\rm initial}$. Furthermore, the off-diagonal panels that plot the degeneracies between the quenching probability parameters and the quenching timescale parameters exhibit expected correlation between the parameters: the longer the quenching timescale the larger the quenching probability correction factor (f_{P_Q}).

We compare the SSFR distribution generated from our model using the median model parameter values of the posterior distribution (orange) to the SSFR distribution of the SDSS DR7 central galaxy catalog (black dashed), in Figure 4.6. The SSFR distribution are computed for four stellar mass bins. We find good agreement between the SSFR distributions in each of the bins. More importantly, the model with parameters values from the posterior distribution is able to successfully reproduce the height of the green valley.

In Figure 4.7, we plot the central galaxy quenching timescale $\tau_Q^{\rm cen}(\mathcal{M})$ corresponding to the median parameter values of the posterior (red points) and compare it to the satellite quenching timescale in Wetzel et al. (2013). We also plot $\tau_Q^{\rm cen}(\mathcal{M})$ for A_τ and δ_τ of the final iteration ABC parameter pool (light red lines) and error bars on median $\tau_Q^{\rm cen}$ to represent the 1-sigma values in stellar mass bins of width $\Delta \log \mathcal{M} = 0.25$ dex. The model used in Wetzel et al. (2013) to infer the satellite quenching timescale has notable difference from our model. However, an analogous analysis reproduces an equivalent satellite quenching timescale. The comparison of the quenching timescales reveal that both timescales exhibit significant mass dependence, which curiously appear to have similar slopes. The similarity, however, is difficulty to precisely quantify because of the uncertainties in both timescales. The comparison, above all, illustrates that *the quenching timescale of central galaxies is significantly longer than the quenching timescale of satellites.*

To determine whether our constraints on the central galaxy timescale are robust, we carry out a similar analysis where we fix the quenching timescale parameters to the satellite quenching timescale of Wetzel et al. (2013). Then we use ABC-PMC with Eq. 4.17 as the distance metric to constrain the parameters $A_{\rm GV}$, $\delta_{\rm GV}$, A_{P_Q}, and δ_{P_Q}. In Figure 4.8, we plot the SSFR distribution generated from median parameter values of the parameter constraints and compare it to the SSFR distribution of the SDSS DR7 central galaxies. At all stellar mass bins, while the quiescent fraction is generally reproduced, the height of the green valley for the model using satellite quenching timescale is significantly lower than the green valley of the SDSS DR7 centrals. Therefore, a longer quenching timescale is necessary to reproduce the height of green valley for central galaxies.

In addition to the quenching timescale constraints, the posterior probability distributions of our model parameters in Figure 4.5, also produce constraints for the quenching probability (Eq. 4.13). Recent works such as Moustakas et al. (2013) and Lian et al. (2016) have published measurements of a comparable quantity: *the quenching rate*. At $0.02 < z < 0.05$ and in three mass bins between 10^{10} and $10^{10.6}$ $M_\odot$, Lian et al. (2016) measures quenching rates of $19, 25,$ and 33%/Gyr. Similarly, Moustakas et al. (2013) measures the quenching rate for four stellar mass between $10^{9.5}$ and $10^{11.5}$ $M_\odot$ out to $z \sim 0.8$. At $z > 0.2$, the Moustakas et al. (2013) quenching rates range between $1 - 12\%$/Gyr. These quenching rates are generally in good agreement with the P_Q from our constraints.

We refrain from a more detailed comparison due to a number of underlying differences between the quenching rates in the literature and our P_Q. For instance, these quenching rates are derived for the entire galaxy population and not the central galaxy population. Furthermore, our P_Q describes the probability that a star-forming central galaxy begins to quench, not the rate at which star-forming galaxies become quiescent. We also note that P_Q is dictated by the SMF and quiescent fraction evolution, so a detailed comparison would require a detailed comparison of the different SMF and quiescent fraction evolutions.

In our model, we obtain stellar masses of central galaxies from the SHAM prescription of host subhalos. As a result, the stellar mass evolution of our central galaxies is sensitive to the SMF and its evolution. In our SHAM procedure, we formulate the SMF based on Li & White (2009) and Marchesini et al. (2009), which evolves significantly for $\mathcal{M} < 10^{11}\mathrm{M}_{\odot}$ over $z < 1$. SMF measurements from PRIMUS for $z < 1$ in Moustakas et al. (2013), however, fail to find such significant SMF evolution. To confirm whether or not our central quenching timescale constraint remains robust over different degrees of SMF evolution, we test our results with two extreme models of SMF evolution (included in Figure 4.1): (1) a model in which the SMF does not evolve with time, and (2) a model in which the SMF at $z = 1.2$ is roughly half of our fiducial SMF at $z = 1.2$. We plot the results in Figure 4.9. We plot the $\tau_Q^{\mathrm{cen}}(\mathcal{M})$ of the median posterior parameter values from our analysis using extreme models of SMF evolution. While the SMF evolution impacts the mass dependence, $\tau_Q^{\mathrm{cen}}(\mathcal{M})$ remains significantly longer than the quenching timescale of satellites.

We also repeat the analysis for different parameterizations of $f_{\mathrm{Q}}^{\mathrm{cen}}$; more specifically, the two $f_{\mathrm{Q}}^{\mathrm{cen}}$ parameterization in Wetzel et al. (2013). Regardless of the $f_{\mathrm{Q}}^{\mathrm{cen}}$ parameterization, we find that $\tau_Q^{\mathrm{cen}}(\mathcal{M})$ is greater than the satellite quenching timescale. We conclude that our central quenching timescale results are robust over the specific choices we make in implementing our model.

4.7 Discussion

4.7.1 Central versus Satellite Quenching

One key result of the central galaxy quenching timescales we infer is its difference with the satellite galaxy quenching timescale from Wetzel et al. (2013). For the entire stellar masses range probed, the quenching timescale of central galaxies is ~ 0.5 Gyr longer than that of satellite galaxies. This corresponds to central galaxies taking approximately ~ 2 Gyrs longer than satellite galaxies to transition from the SFMS to the quiescent peak. Moreover, this difference suggests that *quenching mechanisms responsible for the cessation of star formation in central galaxies are different from the ones in satellite galaxies.*

At a glance, this difference in central and satellite quenching timescale is rather unexpected since the SSFR distribution of central (blue) and satellite (orange) galaxies of the SDSS DR7 Group Catalog in Figure 4.10, show remarkably similar green valley heights. However, the similarity in green valley height is not determined by the quenching timescale alone. It reflects the combination of quenching timescale and the rate that star-forming galaxies transition to quenching. Since the satellite quenching timescale is shorter than that of centrals, star-forming satellites transition to quenching at a higher rate than star-forming centrals at $z = 0$. The difference in this transition rate is even higher than what the quenching timescale reflects because tidal disruption and mergers preferentially destroy quiescent satellite galaxies.

The implication that satellites and centrals have different quenching mechanisms is broadly consistent with the currently favored dichotomy of quenching mechanisms: satellite galaxies undergo environmental quenching while central galaxies undergo internal quenching. It is also consistent with the significant difference in the structural properties of quiescent satellites versus centrals (Woo et al., 2016), which also suggests different physical pathways for quenching satellites versus centrals. Furthermore, it explains the environment dependence in the quiescent fraction evolution in recent observations (Hahn et al. 2015; Darvish et al. 2016). Both central and satellite quenching contribute in high density environments while only central quenching contributes in the field causing the quiescent fraction to increase more significantly in high density environments.

Additionally, combined with the Wetzel et al. (2013) result that at $\mathcal{M}_* > 10^{10}\mathrm{M}_\odot$ central galaxy quenching is the dominant contributor to the growth of the quiescent population, we can also characterize mass regimes where environmental or internal quenching mechanisms dominate, similar to Peng et al. (2010). Below $\mathcal{M}_* < 10^{9}\mathrm{M}_\odot$, satellite quenching is the *only* mechanism (Geha et al., 2012b) and internal quenching is ineffective. Until $\mathcal{M}_* < 10^{10}\mathrm{M}_\odot$, environmental quenching continues to be the dominant mechanism. At $\mathcal{M}_* > 10^{10}\mathrm{M}_\odot$ internal quenching dominates.

4.7.2 Quenching Star Formation in Central Galaxies

Numerous physical processes have been proposed in the literature to explain the quenching of star formation. Observations, however, have yet to identify the primary driver of quenching or consistently narrowing down proposed mechanisms. The quenching timescale we derive for central galaxies provides a key constraint for any of the proposed mechanisms. Only processes that agree with our central galaxy quenching timescales, can be the main driver for quenching star formation in central galaxies.

Merger driven quenching has often been proposed as a driving mechanism of star formation quenching (Springel et al. 2005; Hopkins et al. 2006, 2008a,b). In this proposed mechanism, quenching is typically driven by gas-rich galaxy mergers which induce starburst and rapid black hole growth. Cosmological hydrodynamics simulations that examine mergers, however, conclude that quenching from mergers alone cannot produce a realistic red

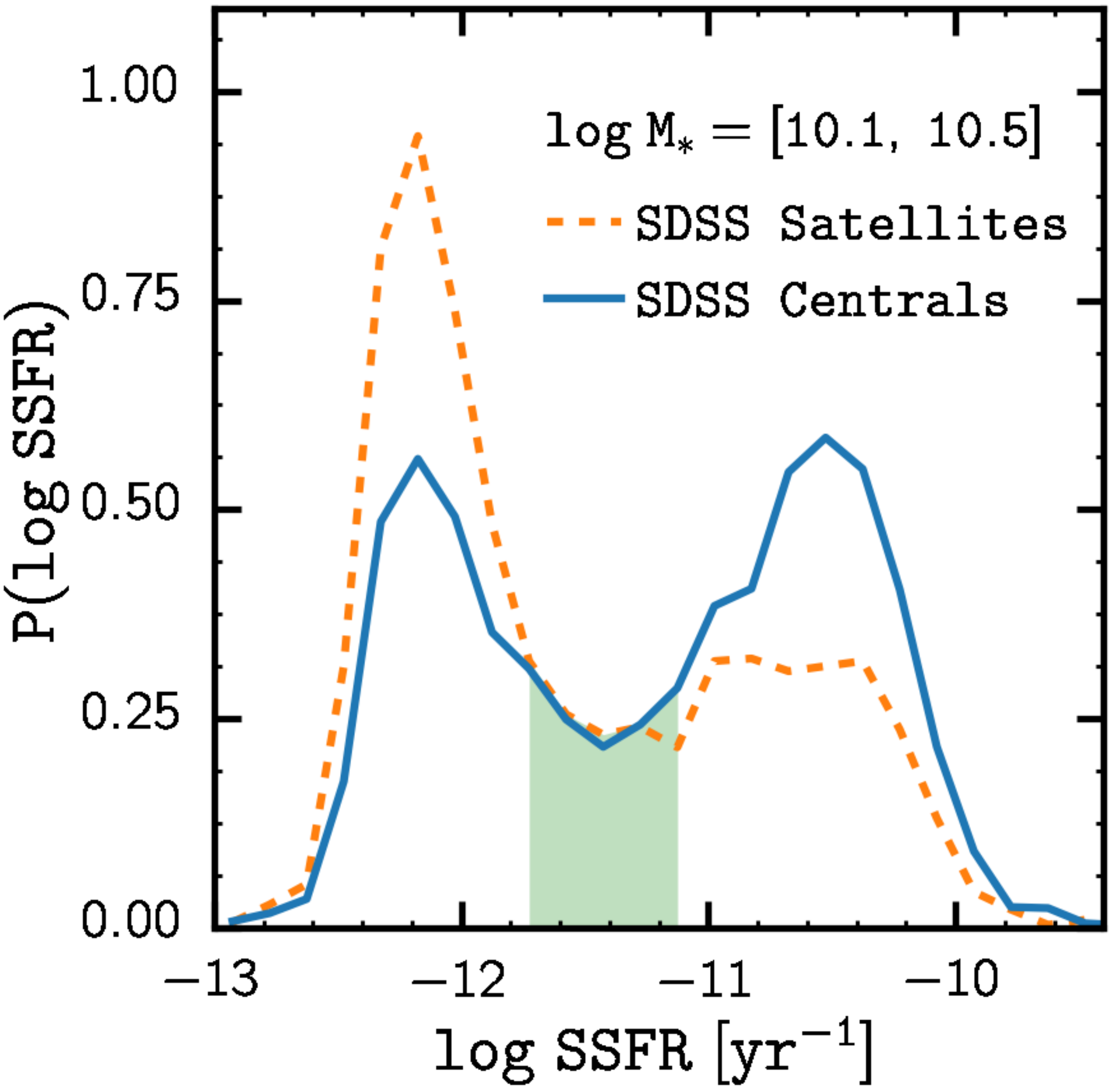

Figure 4.10: SSFR distributions of the central galaxies versus the satellite galaxies in the SDSS DR7 Group Catalog with stellar mass between $10^{10.1}$ and $10^{10.5} M_\odot$. Both SSFR distributions have similar green valley heights (green shaded region). Since central galaxies have significantly longer quenching timescales, satellite galaxies have a higher rate of transitioning from star-forming to quenching than central galaxies.

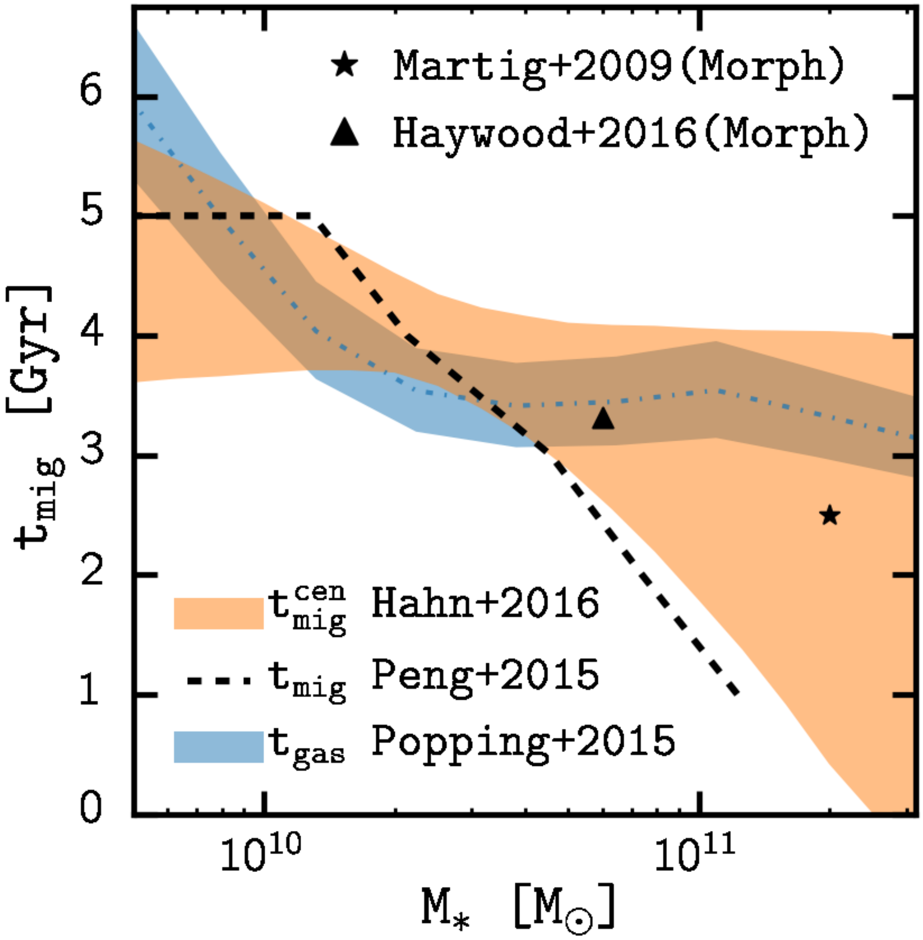

Figure 4.11: Comparison of the central galaxy quenching migration time estimate we infer, ($t_{\rm mig}^{\rm cen}$; orange) with quenching time estimates for gas depletion absent accretion (strangulation) and morphological quenching. The width represents the 68 % confidence region propagated from the posterior distributions of the $\tau_{\rm Q}^{\rm cen}$ parameters. For strangulation, we include the gas depletion time at $z = 0.2$ derived from the star formation efficiency estimates in Popping et al. (2015) (blue dash-dotted). The surrounding blue shaded region plots the range of gas depletion times at $z = 0.15$ (longer) to 0.25 (shorter). We also include the quenching migration time inferred from the Peng et al. (2015) gas regulation model (dashed). For morphological quenching we plot the quenching times taken from the star formation histories of the simulated galaxy in Martig et al. (2009) (star). We also include the quenching times of the Milky Way in Haywood et al. (2016) (triangle). The quenching timescale of strangulation exhibit a similar stellar mass dependence and is generally consistent with our central quenching timescales. Although its feasibility for a wider galaxy population is unexplored, the quenching timescale from morphological quenching is in good agreement with our timescale.

sequence (Gabor et al. 2010, 2011. Gabor et al. (2011) used an on-the-fly prescription to identify mergers and halos in order to test different prescriptions for quenching star formation. In addition to failing to produce a realistic red sequence, they find that mergers cannot sustain quiescence due to gas accretion from the inter-galactic medium, which refuels star formation after $1-2$ Gyr. The major mergers examined in the four high resolution zoom in cosmological hydrodynamic simulation of Sparre & Springel (2016) also fail to sustain quiescence after $1-2$ Gyr (Sparre et al. in prep.).

AGN feedback has also been proposed as a quenching mechanism (Kauffmann & Haehnelt 2000; Croton et al. 2006; Hopkins et al. 2008a; van de Voort et al. 2011), sometimes in conjunction with mergers as a way to sustain quiescence or on its own. The feedback of the AGN deposit sufficient energy, which subsequently prevents additional gas from cooling. A number of more recent works have, however, cast doubt on the role of the AGN in quenching. Mendel et al. (2013), identified quenched galaxies, with selection criteria analogous to the selection of post-starburst galaxies, in the SDSS DR7 sample and found no excess of optical AGN in them, suggesting that AGN do not have defining role in quenching. Gabor & Bournaud (2014) further argue against AGN quenching by examining gas-rich, isolated disk galaxies in a suite of high resolution simulations where they find that the AGN outflows have little impact on the gas reservoir in the galaxy disk and furthermore fail to prevent gas inflow from the intergalactic medium. Yesuf et al. (2014) examined post-starburst galaxies transitioning from the blue cloud to the red sequence to find a significant time delay between the AGN activity and starburst phase, which suggests that AGN do not play a primary role in triggering quenching. AGN may yet be responsible for quenching in conjunction with other mechanisms or have a role in sustaining quiescence.

Besides mergers and AGN driven processes, another class of proposed mechanisms involves some process(es) that restrict the inflow of cold gas – strangulation. With little inflow of cold gas, the galaxy quenches as it depletes its cold gas reservoir. One mechanism that has been proposed to prevent cold gas accretion is loosely referred to as "halo quenching". A hot gaseous coronae, which form in halos with masses above $\sim 10^{12}\mathrm{M}_\odot$ via virial shocks, starves galaxies of cold gas for star formation (Birnboim & Dekel 2003; Kereš et al. 2005; Cattaneo et al. 2006; Dekel & Birnboim 2006; Birnboim et al. 2007; Gabor & Davé 2012, 2015). For these sorts of mechanisms, the quenching timescale is linked to the time it takes for the galaxy to deplete its cold gas reservoir – the gas depletion timescale.

In principle, the gas depletion time can be estimated from measurements of the total gas mass or gas fraction. In Popping et al. (2015), for instance, they derive "star formation efficiency" (SFE; inverse of the gas depletion time) by dividing the SFR of the SFMS by the total galaxy gas mass that they infer from their semi-empirical model. These sorts of gas depletion time estimates, however, have significant redshift dependence because the gas fraction of galaxies do not evolve significantly over $z < 1$ (Stewart et al. 2009; Santini et al. 2014; Popping et al. 2015).

Nevertheless, in Figure 4.11 we estimate the central quenching migration time ($t^\mathrm{cen}_\mathrm{mig}$; orange) – the time it takes central galaxies to migrate from the SFMS to quiescent estimated from our $\tau^\mathrm{cen}_\mathrm{Q}$ – to the gas depletion times derived from the Popping et al. (2015) SFEs (blue). For $t^\mathrm{cen}_\mathrm{mig}$, we compute the time it takes a quenching galaxy to transition from the SFMS to the quiescent peak of the SFR distribution at $z = 0.2$. We compute $t^\mathrm{cen}_\mathrm{mig}$ at $z = 0.2$ because this is approximately when the $z \approx 0$ green valley galaxies would have started quenching. For the gas depletion time, we invert the SFE at $z = 0.2$, interpolated between the $z = 0.$ and $z = 0.5$ Popping et al. (2015) SFEs (blue dot-dashed). The surrounding blue shaded region marks the range of gas depletion times from $z = 0.15$ (longer) to 0.25 (shorter) to illustrate the significant redshift dependence. We also note that over the redshift range $z = 0.5$ to 0., at $\mathcal{M} = 10^{10}\mathrm{M}_\odot$, the Popping et al. (2015) gas depletion time varies from ~ 2.5 to 7 Gyrs.

The $t^{\rm cen}_{\rm mig}$ and gas depletion time in Figure 4.11 are generally in agreement with each other and exhibit similar mass dependence.

Beyond the estimates of gas depletion times from gas mass, recently Peng et al. (2015), using a gas regulation model (e.g. Lilly et al. 2013; Peng & Maiolino 2014), explored the impact that different quenching mechanisms have on the stellar metallicity of local galaxies from the SDSS DR7 sample. To reproduce the stellar metallicity difference between quiescent and star forming galaxies in their galaxy sample, they conclude that the primary mechanism for quenching is gas depletion absent accretion and it has a typical quenching migration time of $t_{\rm mig} \sim 4$ Gyr for $\mathcal{M} < 10^{11} \mathrm{M}_\odot$. We infer the quenching migration time from Figure 2 of Peng et al. (2015) and include it in Figure 4.11 (dashed). The Peng et al. (2015) migration time exhibits a similar mass dependence as our central quenching migration time. Furthermore, although slightly shorter at $\mathcal{M} > 5 \times 10^{10} \mathrm{M}_\odot$, the migration time is broadly consistent with our central quenching migration time.

Overall, our $t^{\rm cen}_{\rm mig}$ is consistent with the migration time estimates of gas depletion mechanisms. In other words, our central galaxy quenching timescale is consistent with the timescales predicted by gas depletion absent accretion. One currently favored model for halting cold gas accretion – halo quenching – quenches galaxies that inhabit host halos with masses greater than some threshold $\sim 10^{12} \mathrm{M}_\odot$. Based on SHAM, this halo mass threshold corresponds to stellar masses of $\sim 10^{10.25} \mathrm{M}_\odot$. Yet, a significant fraction of the SDSS central galaxy population with stellar masses $< 10^{10.25} \mathrm{M}_\odot$ are quiescent. While, scatter in the halo mass threshold and the stellar mass to halo mass relation, combined, may help resolve this tension, halo quenching faces a number of other challenges. For instance, the predictions of halo quenching models are difficult to reconcile with the observed scatter in the stellar mass to halo mass relation (Tinker 2016). Furthermore, models that rely only on such halo quenching still must account for the hot gas in the inner region of the halo, which, because of its high density, often has short cooling times of just $1-2$ Gyr. Of course, the challenges of halo quenching does *not* rule out quenching from gas depletion absent accretion since other mechanisms may also prevent cold gas from accreting onto the central galaxy

Finally, morphological quenching has also been proposed as a mechanism responsible for quenching star formation. In the mechanism proposed by Martig et al. (2009), for instance, star formation in galactic disks are quenched once the galactic disks become dominated by a stellar bulge. This stabilizes the disk from fragmenting into bound, star forming clumps. In a cosmological zoom-in simulation of a $\sim 2 \times 10^{11}\mathrm{M}_\odot$ galaxy selected to examine such a mechanism, Martig et al. (2009) finds that the galaxy quenches its star formation from $\sim 10\ \mathrm{M}_\odot\mathrm{yr}^{-1}$ to $\sim 1.5\ \mathrm{M}_\odot\mathrm{yr}^{-1}$ in ~ 2.5 Gyr during the morphological quenching phase. A $\mathcal{M} \sim 2 \times 10^{11}\mathrm{M}_\odot$ galaxy with $\hat{t}_\mathrm{Q} \sim 2.5$ Gyr (star; Figure 4.11) is in good agreement with $\hat{t}_\mathrm{Q}^\mathrm{cen}$. Despite this agreement, morphological quenching faces a number of challenges. There is little evidence from modern cosmological hydrodynamic simulations that suggest that morphological quenching can drive anything beyond short timescale fluctuations in gas fueling and SFR. Furthermore, proposed morphological quenching mechanisms face the "cooling flow problem" where they fail to prevent gas cooling onto a galaxy. Without addressing this issue, proposed morphological quenching mechanisms *cannot* maintain quiescence.

Our own Milky Way galaxy, as Haywood et al. (2016) finds, after forming its bar undergoes quenching. In the star formation history of the Milky Way that Haywood et al. (2016) recovers, the SFR of the Milky Way decreases by an order of magnitude over the span of roughly 1.5 Gyr. Converting to $\hat{t}_\mathrm{Q}$ in a similar fashion as our $\hat{t}_\mathrm{Q}^\mathrm{cen}$ estimates and assuming a Milky Way stellar mass of $\sim 6 \times 10^{10}\mathrm{M}_\odot$ (Licquia & Newman 2015; Haywood et al. 2016), we find remarkable agreement with our $\hat{t}_\mathrm{Q}^\mathrm{cen}$ (Figure 4.11). Motivated by the contemporaneous formation of the bar with quenching, Haywood et al. (2016) suggest a bar driven (morphological) quenching mechanism that inhibits gas accretion through high level turbulence supported pressure that is generated from the shearing of the gaseous disk. Although, this proposal may resolve the cooling-flow problem, their arguments for the mechanism are qualitative and thus require more detailed investigation. Admittedly, however, this particular comparison is hastily made since quenching event occurs beyond the redshift probed by our simulation at $1 < z < 2$. Furthermore, after dramatic quenching episode, based on the star formation history that Haywood et al. (2016) recovers, the Milky Way resumes star formation at a much lower level.

The central quenching timescale we infer from our analysis provides key insight into the physical processes responsible for quenching star formation. It offers a means of assessing the feasibility of numerous quenching mechanisms, which operate on distinct timescale. Based on the latest models and simulations, merger driven quenching has fallen out of favor and AGN alone seem insufficient in triggering quenching. Mechanisms that halt cold gas accretion, such as halo quenching, predict quenching times generally consistent with our estimates from the central quenching timescale we derive. However, it fails to explain the significant low mass quiescent population of central galaxies. Morphological quenching, with its agreement in quenching time, may be a key physical mechanism in quenching star formation. However, more evidence is required that it can address the cooling flow problem and maintain quiescence. Furthermore, its role in the overall quenching of galaxy populations – not just single simulated galaxies – still remains to be explored.

4.8 Summary

Understanding the physical mechanisms responsible for quenching star formation in galaxies has been a long standing challenge for hierarchical galaxy formation models. Following the success of Wetzel et al. (2013) in constraining the quenching timescales of satellite galaxies, in this work, we focus on star formation quenching in central galaxies with a similar approach. Using a high resolution N-body simulation in conjunction with observations of the SMF, SFMS, and quiescent fraction at $z < 1$, we construct a model that statistically tracks the star formation histories of central galaxies. The free parameters of our model dictate the height of the green valley at the initial redshift, the correction to the quenching probability, and most importantly, the quenching timescale of central galaxies,

Using ABC-PMC with our model, we infer parameter constraints that best reproduce the observations of the central galaxy SSFR distribution from the SDSS DR7 Group Catalog and the central galaxy quiescent fraction evolution. From the parameter constraints of our model, we find the following results:

1. The quenching timescale of central galaxies exhibit a significant mass dependence: more massive central galaxies have shorter quenching timescales. Over the stellar mass range $\mathcal{M} = 10^{9.5} - 10^{11.5}\mathrm{M}_{\odot}$, $\tau_{\mathrm{Q}}^{\mathrm{cen}} \sim 1.2 - 0.5$ Gyr. Based on these timescales, central galaxies take roughly 2 to 5 Gyrs to traverse the green valley.

2. The quenching timescale of central galaxies is significantly longer than the quenching timescale of satellite galaxies. This result is robust for extreme prescriptions of the SMF evolution in our simulation and even for different parameterizations of the central quiescent fraction.

3. The difference in quenching timescales of satellite and centrals suggest that different physical mechanisms are primary drivers of star formation quenching in satellites versus centrals. Satellite galaxies experience external "environment quenching" while central galaxies experience internal "self quenching".

4. We compare the central quenching timescales we infer to the gas depletion timescales predicted by quenching through strangulation and find broad agreement. We also find

good agreement with morphological quenching; however, its feasibility in maintaining quiescent and for a wider galaxy population remains to be explored.

Ultimately, the central galaxy quenching timescale we obtain in our analysis provides a crucial constraint for any proposed mechanism for star formation quenching.

One key component of our simulation is the use of SHAM to track evolution of stellar masses of central galaxies. As mentioned above, the central galaxy quenching timescale results we obtain remain unchanged if we use stellar mass growth from integrated SFR. However, the use of SHAM stellar masses neglects the connection between stellar mass growth and star formation history. To incorporate integrated SFR galaxy stellar mass growth in our simulation, however, a better understanding of the detailed relationship among stellar mass growth, host halo growth, and the observed stellar mass to halo mass relation is required. We will explore this in future work.

Acknowledgements

CHH was supported by NSF-AST-1109432 and NSF-AST-1211644. AW was supported by a Caltech-Carnegie Fellowship, in part through the Moore Center for Theoretical Cosmology and Physics at Caltech. We thank Charlie Conroy and Michael R. Blanton for helpful discussions. We also thank the referee for the thorough review of our work and the constructive report that substantially improved the quality of the paper. CHH also thanks the Instituto de Física Teoórica (UAM/CSIC) and particularly Francisco Prada for their hospitality during his summer visit, where part of this work was completed.

Conclusion

Over the next decade future surveys, namely eBOSS and DESI, will expand the cosmic volumes probed with redshifts by an order of magnitude. They have the potential to measure the growth of structure and constrain cosmological parameters with unprecedented precision. The main challenges for realizing their full statistical power are methodological. The frameworks I present in this dissertation – robust treatment of systematics, innovative approaches to accurate inference, and improved models of the galaxy-halo connection – can be extended to these future surveys and used to tackle key methodological challenges.

For instance, observational systematics such as fiber collisions will continue to impact galaxy clusteing analyses of eBOSS and DESI, which will utilize fiber-fed spectrographs. As described in Chapter 1, due to the impact that fiber collisions have on small scales, much of the statistical gains from eBOSS and DESI will be *wasted* if they are not properly account for in analyses. In fact, in eBOSS and DESI the systematics will be more complicated with multiple classes of target objects and automated fiber positioning (Cahn et al., 2015; Dawson et al., 2015). But the methods from Chapter 1 can be extended to both surveys.

Furthermore, in Chapter 2, we revealed deviations between the ABC posterior probability distribution and the standard Gaussian pseudo-likelihood approach to inference – even in the narrower context of halo occupation modeling. Yet there have not been direct investigations on the impact of the standard assumptions on more general cosmological parameter constraints. With the increased statistical power of future surveys, quantifying the impact of these assumptions in our inference is critical for unbiased constraints. While tractability of forward modeling the data has been an obstacle for adopting ABC, new models aimed at the next galaxy surveys, are making promising strides.

Finally, as we describe in Chapter 3, observations of galaxies have firmly established a global view of galaxy properties out to $z{\sim}1$. As in Chapter 4, precise predictions of hierarchical growth of structure from ΛCDM can be used to constrain key elements of galaxy evolution in a data-driven and statistical fashion. The introduction of Integral Field Unit observations (*e.g.* MaNGA) and larger galaxy samples (*e.g.* DESI Bright Galaxy Survey) offer exciting opportunities to extend the works of Chapters 3 and 4 and construct better models of the galaxy-halo connection.

Each aspect of my dissertation will be instrumental for exploiting the full potential of future surveys and making more precise measurements of the growth rate of structure, the cosmological parameters, and thus tests of General Relativity and modified gravity scenarios. Furthermore, galaxy clustering also provides a unique window to probe fundamental physics — *i.e.* the total neutrino mass (Σm_ν). Extending the methods from my dissertation to future surveys will allow us to better measure the imprints of neutrinos on LSS and produce tigher constraints on Σm_ν. A tighter upper limit on Σm_ν is essential to distinguish between the neutrino mass hierarchies and will provide an important input for particle physics theory beyond the Standard Model.

Appendix A

Effective Window Method Polynomials

For reference, here we list the first few polynomials $H_{l_> l_<}(x)$ from Eq. (1.33)

$$H_{20}(x) = x^2 - 1, \tag{20}$$

$$H_{40}(x) = \frac{7}{4}x^4 - \frac{5}{2}x^2 + \frac{3}{4}, \tag{21}$$

$$H_{42}(x) = x^4 - x^2, \tag{22}$$

$$H_{60}(x) = \frac{33}{8}x^6 - \frac{63}{8}x^4 + \frac{35}{8}x^2 - \frac{5}{8}, \tag{23}$$

$$H_{62}(x) = \frac{11}{4}x^6 - \frac{9}{2}x^4 + \frac{7}{4}x^2, \tag{24}$$

$$H_{64}(x) = x^6 - x^4 \tag{25}$$

Appendix B

Indicators of Star Formation

In order to measure star formation in galaxies of the SDSS DR7 group catalog, we use specific star formation rates (SSFR) derived in Brinchmann et al. (2004) (briefly described in § 4.3). These SSFR are measured from Hα emission lines, and D_n4000 for SSFRs$\gtrsim 10^{-12}\mathrm{yr}^{-1}$. While no SFR indicator provides the panacea for uncertainties in measuring star formation in galaxies, a number of caveats must be addressed for the Brinchmann et al. (2004) SSFR. SSFRs derived from Hα probe star formation on a $\sim$ 10 Myr timescale, which makes them sensitive to short term varations in the galaxies' star formation histories (Kennicutt & Evans, 2012). Furthermore, the spectroscopically derived Brinchmann et al. (2004) SSFRs also rely on aperture corrections, which may introduce further uncertainties. In this section, we demonstrate, by comparing to another SFR indicator, that our specific choice of SFR indicator does not significant impact the central galaxy quenching timescale.

In Moustakas et al. (2013) (hereafter M2013), for their lowest redshift galaxy sample, they construct a catalog derived from the SDSS DR7 VAGC. They supplement the optical photometry from SDSS DR7 with UV photometry from GALEX, integrated J H K_s magnitudes from 2MASS Extended Source Catalog, and integrated photometry at 3.4 and 4.6μm

from the WISE All-Sky Data Release[2]. Then to derive galaxy properties such as $\mathcal{M}_*$ and SFR, they use `iSEDfit` – a Bayesian SED modeling code. By including UV photometry from GALEX, the SFRs from the M2013 catalog traces star formation over $\sim 10 - 100$ Myr timescales and is not dominated by short term variations. Furthermore, as these SFRs are derived from photometry, they do not require any aperture correction.

Galaxies that are in both the SDSS DR7 group catalog and M2013 catalog, provide a convenient galaxies sample to compare the distinct SFR indicators. In Figure 12, we compare the SSFR distributions of this subsample, calculated using SSFRs from the SDSS DR7 group catalog (black dashed) versus M2013 (orange): $P(\mathrm{SSFR}^{\mathrm{group}})$ versus $P(\mathrm{SSFR}^{M2013})$. Before comparing the $P(\mathrm{SSFR})$s, we note that the SSFRs from M2013 are not subject to the Brinchmann et al. (2004) SSFR upper bounds for low star-forming galaxies (see § 4.3). That is, the M2013 SSFRs can extend below 10^{-13} yr^{-1}. For a meaningful comparison, however, we impose similar SSFR bounds to reproduce the $P(\mathrm{SSFR}^{\mathrm{group}})$ quiescent peak. We also note that due to the M2013 bright magnitude limit, the M2013 sample does not contain a large number of galaxies within the group catalog's z range at higher mass bins. Furthermore, for both distributions, the galaxies are binned based on the group catalog $\mathcal{M}_*$ so that the same galaxies are examined in each bin. This binning does not have a significant impact on the comparision because the group catalog $\mathcal{M}_*$ and M2013 $\mathcal{M}_*$ are tightly correlated.

There are some minor discrepancies between the SSFR distributions, such as the position of the star-forming peak in the lowest mass bin. While this is caused by small differences in the slopes of the SFMS between the M2013 sample and the group catalog, the star-forming peaks in the higher mass bins are in good agreement. So for the $\mathcal{M}_*$ probed by our analysis, this discrepancy does not have a significant impact. Overall, however, the $P(\mathrm{SSFR})$s are in good agreement with one another. Furthermore, we find that the heights of

[2] http://wise2.ipac.caltech.edu/docs/release/allsky

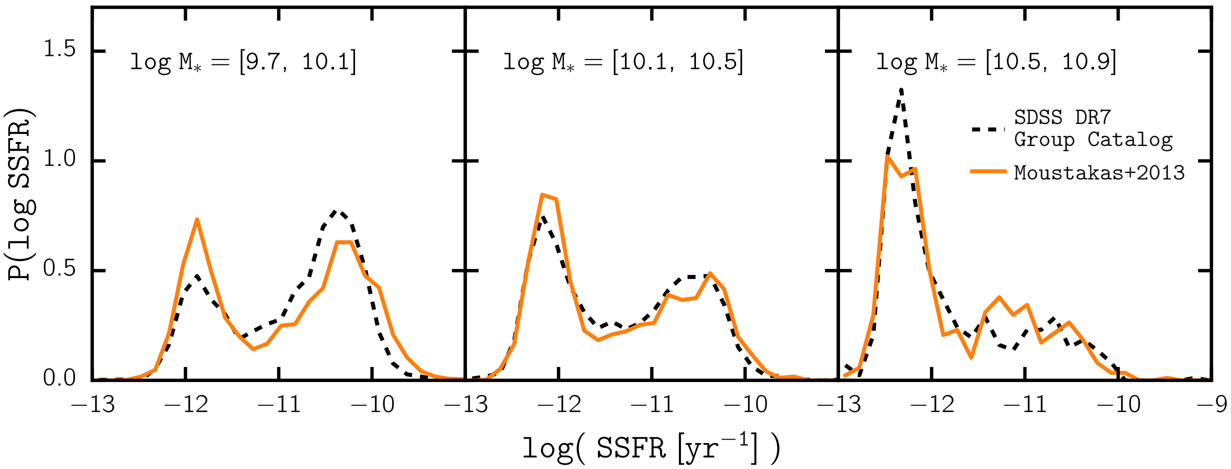

Figure 12: Comparison of the SSFR distribution calculated using SSFRs from the SDSS DR7 group catalog (black dashed) versus Moustakas et al. (2013) (orange) for galaxies that are in both the SDSS DR7 group catalog and the Moustakas et al. (2013) sample $z \sim 0.1$ bin: $P(\mathrm{SSFR}^{\mathrm{group}})$ versus $P(\mathrm{SSFR}^{M2013})$. Galaxies are binned based on the group catalog $\mathcal{M}_*$ for both distributions so that the same galaxies are examined in each bin. We impose SSFR bounds on $P(\mathrm{SSFR}^{M2013})$ for low SSFRs to reproduce the $P(\mathrm{SSFR}^{\mathrm{group}})$ quiescent peak (§ 4.3). We note that the M2013 sample does not contain a large number of galaxies within the group catalog's z range at higher mass bins due its bright magnitude limit. We find good overall agremeent between $P(\mathrm{SSFR}^{\mathrm{group}})$ and $P(\mathrm{SSFR}^{M2013})$. Furthermore, they have consistent green valley heights, which is the main feature of $P(SSFR)$ critical for constraining the central quenching timescale.

the green valley in both distributions, the main feature of $P(SSFR)$ critical for constraining the central quenching timescale, are also in good agreement. Therefore, we conclude that the Brinchmann et al. (2004) SFRs do not significantly impact the quenching timescale and the results of this work.

Appendix C

Star-forming/Quiescent Classifications

In our parameterization of the observed $f_{\rm Q}^{\rm cen}$ in Eqs. 4.1 and 4.2, we derive the best-fit values for the parameters A_0 and A_1 by fitting $f_{\rm Q}^{\rm cen}$ measured in the SDSS DR7 group catalog. The SDSS DR7 group catalog $f_{\rm Q}^{\rm cen}$ is derived using a $\mathrm{SFR} - \mathcal{M}_*$ cut specified in Eq. 4.18. For $\alpha(\mathcal{M}_*)$, the parameter that dictates the $f_{\rm Q}^{\rm cen}$ redshift dependence, however, the best-fit value is derived from fitting Tinker et al. (2013) $f_{\rm Q}^{\rm cen}$ measurements of the COSMOS survey. These $f_{\rm Q}^{\rm cen}$ measurements use $(NUV - R) - (R - J)$ color-color cuts described in Bundy et al. (2010) for the star-forming/quiescent classification. In this section, we demonstrate the consistency between the SDSS DR7 group catalog $f_{\rm Q}^{\rm cen}$, using a $\mathrm{SFR} - \mathcal{M}_*$ cut, and the Tinker et al. (2013) $f_{\rm Q}^{\rm cen}$, using a $(NUV - R) - (R - J)$ color-color cut.

For the galaxies in our SDSS DR7 group catalog, we construct a catalog with UV, optical, and infrared photometry. For UV and optical, we obtain GALEX and SDSS photometry from the NASA-Sloan Atlas[3]. For infrared, we use photometry from the 2MASS all-sky map (Cutri, 2000). We then determine the $FUV, NUV,\ u, g, r, i, z,\ J, H, K_s$ band

[3] http://www.nsatlas.org/

K-corrections and absolute magnitudes for the galaxies using $\texttt{K}-\texttt{correct}$[4] (v4.2 Blanton & Roweis, 2007).

Using these absolute magntidues, in Figure 13, we plot $(NUV - R) - (R - J)$ color-color relation for the SDSS DR7 group catalog (black). We highlight the galaxies in the sample that are classified as quiescent using the $\mathrm{SFR} - \mathcal{M}_*$ cut in orange. Furthermore, we plot the color-color cuts from Bundy et al. (2010) (blue dash-dotted and red dashed lines). Galaxies that lie above both color-color cuts, are classified as quiescent in the $(NUV - R) - (R - J)$ classification.

The horizontal color-color cut (blue dash-dotted) is evaluated using the Bundy et al. (2010) parameterization, at $z \sim 0.0$. The diagonal cut (red dashed) in Bundy et al. (2010) is, however, parameterized using coefficients determined by inspection of redshift bins. Therefore, for the SDSS DR7 group catalog, we extrapolate the coefficients from the COSMOS $z \sim 0.3, 0.7$ bins. We note that using the coefficients from the lowest COSMOS redshift bin ($z \sim 0.3$) instead of extrapolating to $z \sim 0.0$, does not significantly impact the comparison in this section.

Comparison of the quiescent galaxies classified with $\mathrm{SFR}-\mathcal{M}_*$ with the color-color cuts in Figure 13 find that the two classifications are generally consistent. To further test whether the different classifications can impact quiescent fraction parameterization, in Figure 14, we compare the the quiescent fractions derived from them for the SDSS DR7 group catalog: $f_\mathrm{Q}^{SFR-\mathcal{M}_*}$ (black) versus $f_\mathrm{Q}^\mathrm{color}$ (orange). Throughout the mass range of the catalog, $f_\mathrm{Q}^{SFR-\mathcal{M}_*}$ and $f_\mathrm{Q}^\mathrm{color}$ are consistent with each other. Therefore, the $f_\mathrm{Q}^\mathrm{cen}(\mathcal{M}_*, z)$ parameterization derived from measurements of SDSS DR 7 group catalog and Tinker et al. (2013) (Eq. 4.1) does *not* affect the results of this work.

[4] http://howdy.physics.nyu.edu/index.php/Kcorrect

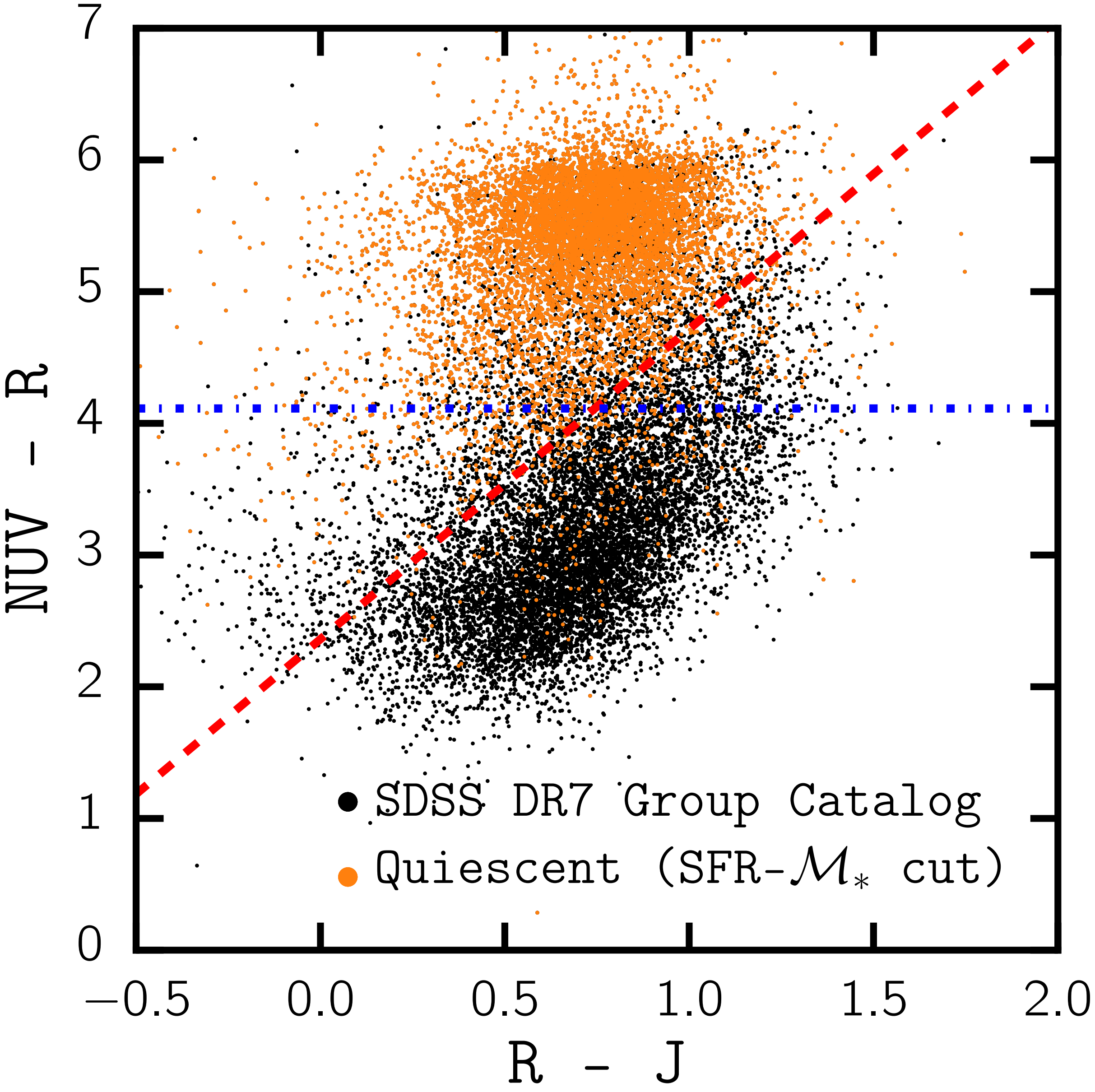

Figure 13: The $(NUV - R) - (R - J)$ color-color relation for the SDSS DR7 group catalog (black) calculated from photometry compiled from GALEX, SDSS, and 2MASS (§ 4.8). Galaxies classified as quiescent using the SFR $- \mathcal{M}_*$ cut are highlighted (orange). Furthermore, we plot the color-color cuts from Bundy et al. (2010) that describe the classification of star-forming/quiescent galaxies in Tinker et al. (2013). We note that the quiescent galaxies classified using the SFR $- \mathcal{M}_*$ cut are generally consistent with the Bundy et al. (2010) color-color cuts.

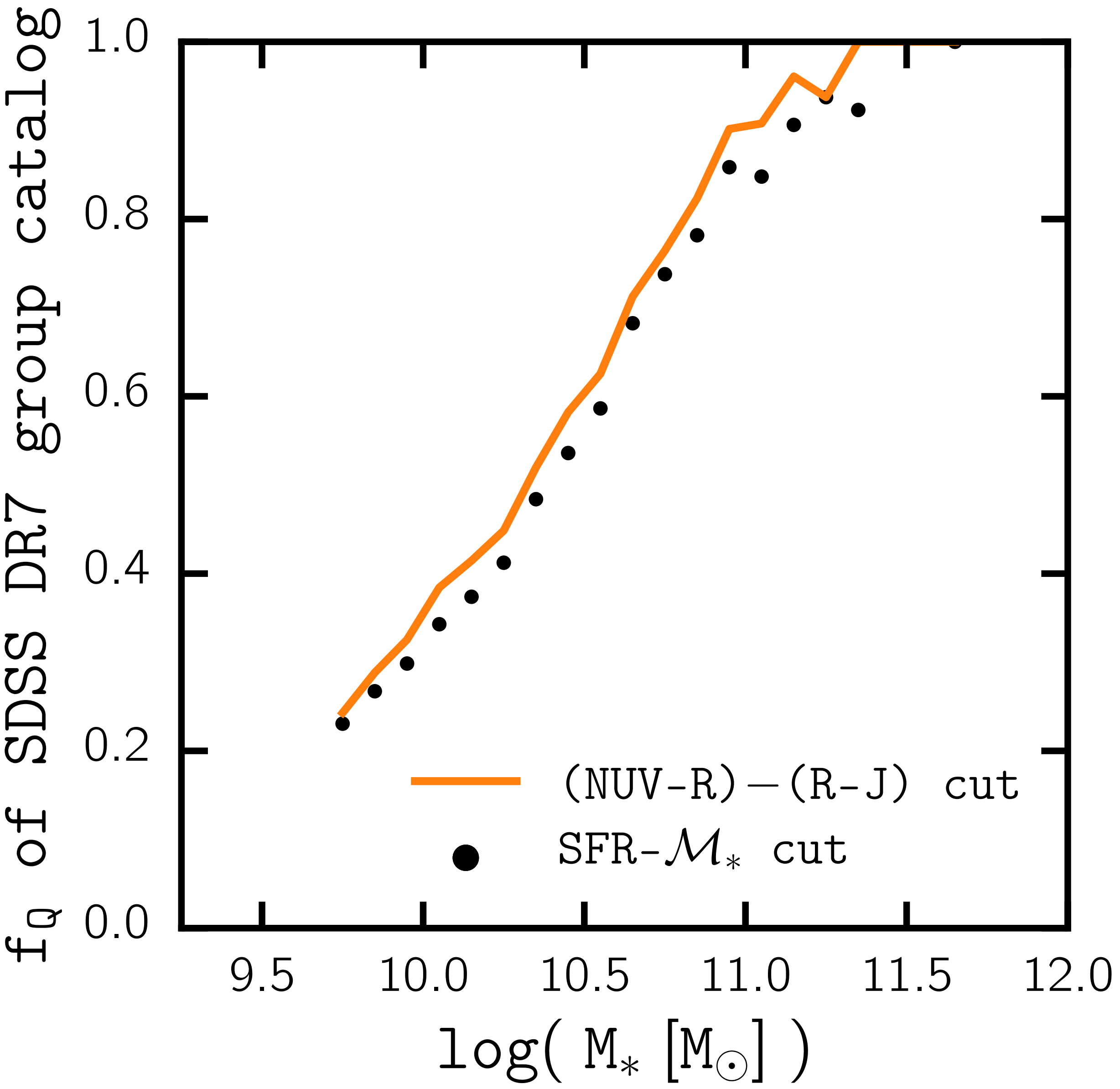

Figure 14: Comparison of the SDSS DR7 group catalog $f_Q(\mathcal{M}_*)$ measured using the SFR − $\mathcal{M}_*$ versus the $(NUV - R) - (R - J)$ color-color classifications. The f_Qs measured using the two different classification methods are consistent with each other. This consistency illustrates that the $f_Q^{\rm cen}(\mathcal{M}_*, z)$ parameterization derived from measurements of SDSS DR 7 group catalog and Tinker et al. (2013) does *not* affect the results of this work.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Akeret, J., Refregier, A., Amara, A., Seehars, S., & Hasner, C. 2015a, J. Cosmology Astropart. Phys., 8, 043

—. 2015b, Journal of Cosmology and Astroparticle Physics, 8, 043

Alam, S., Ho, S., Vargas-Magaña, M., & Schneider, D. P. 2015, MNRAS, 453, 1754

Alam, S., Ata, M., Bailey, S., et al. 2016, ArXiv e-prints, arXiv:1607.03155

Alonso, D. 2012, ArXiv e-prints, arXiv:1210.1833

Anderson, L., Aubourg, E., Bailey, S., et al. 2012, MNRAS, 427, 3435

Ata, M., Kitaura, F.-S., & Müller, V. 2015, MNRAS, 446, 4250

Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, MNRAS, 373, 469

Baldry, I. K., Glazebrook, K., & Driver, S. P. 2008, MNRAS, 388, 945

Balogh, M. L., Navarro, J. F., & Morris, S. L. 2000, ApJ, 540, 113

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., & Robert, C. P. 2009, Biometrika, asp052

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

—. 2013b, ApJ, 770, 57

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

—. 2013d, ApJ, 762, 109

Bekki, K. 2009, MNRAS, 399, 2221

Bell, E. F., Wolf, C., Meisenheimer, K., et al. 2004, ApJ, 608, 752

Bell, E. F., Papovich, C., Wolf, C., et al. 2005, ApJ, 625, 23

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

Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006a, ApJS, 167, 1

—. 2006b, ApJS, 167, 1

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Phys. Rep., 367, 1

Bernardi, M., Hyde, J. B., Sheth, R. K., Miller, C. J., & Nichol, R. C. 2007, AJ, 133, 1741

Beutler, F., Saito, S., Brownstein, J. R., et al. 2014a, MNRAS, 444, 3501

Beutler, F., Saito, S., Seo, H.-J., et al. 2014b, MNRAS, 443, 1065

Beutler, F., Seo, H.-J., Saito, S., et al. 2016, ArXiv e-prints, arXiv:1607.03150

Bianchi, D., Gil-Marín, H., Ruggeri, R., & Percival, W. J. 2015, MNRAS, 453, L11

Bird, S., Viel, M., & Haehnelt, M. G. 2012, MNRAS, 420, 2551

Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349

Birnboim, Y., Dekel, A., & Neistein, E. 2007, MNRAS, 380, 339

Bishop, C. 2007, Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn

Blanton, M. R. 2006, ApJ, 648, 268

Blanton, M. R., & Berlind, A. A. 2007, ApJ, 664, 791

Blanton, M. R., Eisenstein, D., Hogg, D. W., & Zehavi, I. 2006, ApJ, 645, 977

Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, AJ, 142, 31

Blanton, M. R., Lupton, R. H., Schlegel, D. J., et al. 2005a, ApJ, 631, 208

Blanton, M. R., & Moustakas, J. 2009a, ARA&A, 47, 159

—. 2009b, ARA&A, 47, 159

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, ApJ, 594, 186

Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005b, AJ, 129, 2562

Bolzonella, M., Kovač, K., Pozzetti, L., et al. 2010, A&A, 524, A76

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

Borch, A., Meisenheimer, K., Bell, E. F., et al. 2006a, A&A, 453, 869

—. 2006b, A&A, 453, 869

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151

Bundy, K., Ellis, R. S., Conselice, C. J., et al. 2006a, ApJ, 651, 120

—. 2006b, ApJ, 651, 120

Bundy, K., Scarlata, C., Carollo, C. M., et al. 2010, ApJ, 719, 1969

Butcher, H., & Oemler, Jr., A. 1984, ApJ, 285, 426

Cacciato, M., van den Bosch, F. C., More, S., Mo, H., & Yang, X. 2013, MNRAS, 430, 767

Cahn, R. N., Bailey, S. J., Dawson, K. S., et al. 2015, in American Astronomical Society Meeting Abstracts, Vol. 225, American Astronomical Society Meeting Abstracts, 336.10

Cameron, E., & Pettitt, A. N. 2012a, MNRAS, 425, 44

—. 2012b, MNRAS, 425, 44

Campbell, D., van den Bosch, F. C., Hearin, A., et al. 2015, MNRAS, 452, 444

Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 646

Casas-Miranda, R., Mo, H. J., Sheth, R. K., & Boerner, G. 2002, MNRAS, 333, 730

Cassata, P., Cimatti, A., Kurk, J., et al. 2008, A&A, 483, L39

Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J. 2006, MNRAS, 370, 1651

Chabrier, G. 2003, PASP, 115, 763

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

Chuang, C.-H., Zhao, C., Prada, F., et al. 2015a, MNRAS, 452, 686

—. 2015b, MNRAS, 452, 686

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8

Cole, S., Hatton, S., Weinberg, D. H., & Frenk, C. S. 1998, MNRAS, 300, 945

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

Colless, M. 1999, Royal Society of London Philosophical Transactions Series A, 357, 105

Conroy, C., & Gunn, J. E. 2010, FSPS: Flexible Stellar Population Synthesis, astrophysics Source Code Library, ascl:1010.043

Conroy, C., & Wechsler, R. H. 2009a, ApJ, 696, 620

—. 2009b, ApJ, 696, 620

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

Cool, R. J., Moustakas, J., Blanton, M. R., et al. 2013, ApJ, 767, 118

Cooper, M. C., Newman, J. A., Madgwick, D. S., et al. 2005, ApJ, 634, 833

Cooper, M. C., Newman, J. A., Coil, A. L., et al. 2007, MNRAS, 376, 1445

Cooper, M. C., Newman, J. A., Weiner, B. J., et al. 2008, MNRAS, 383, 1058

Cooper, M. C., Coil, A. L., Gerke, B. F., et al. 2010, MNRAS, 409, 337

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

Costanzi, M., Villaescusa-Navarro, F., Viel, M., et al. 2013, J. Cosmology Astropart. Phys., 12, 012

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Cucciati, O., Iovino, A., Kovač, K., et al. 2010, A&A, 524, A2

Cuesta, A. J., Niro, V., & Verde, L. 2016a, Physics of the Dark Universe, 13, 77

Cuesta, A. J., Vargas-Magaña, M., Beutler, F., et al. 2016b, MNRAS, 457, 1770

Cutri, R. M., e. a. 2000

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156

Darvish, B., Mobasher, B., Sobral, D., et al. 2016, ArXiv e-prints, arXiv:1605.03182

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

—. 1985b, ApJ, 292, 371

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013a, AJ, 145, 10

—. 2013b, AJ, 145, 10

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2015, ArXiv e-prints, arXiv:1508.04473

de Putter, R., Wagner, C., Mena, O., Verde, L., & Percival, W. J. 2012, J. Cosmology Astropart. Phys., 4, 019

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2

—. 2008, MNRAS, 383, 119

Del Moral, P., Doucet, A., & Jasra, A. 2006, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68, 411

Desai, V., Dalcanton, J. J., Aragón-Salamanca, A., et al. 2007, ApJ, 660, 1151

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

Dodelson, S. 2003, Modern cosmology

Dressler, A. 1980, ApJ, 236, 351

Dressler, A. 1980, ApJ, 236, 351

Dressler, A. 1984, ARA&A, 22, 185

Drory, N., Bundy, K., Leauthaud, A., et al. 2009, ApJ, 707, 1595

Dutton, A. A., & Macciò, A. V. 2014, MNRAS, 441, 3359

Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605

—. 1999, ApJ, 511, 5

Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, AJ, 122, 2267

Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33

Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227

Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, ApJ, 665, 265

Fasano, G., Poggianti, B. M., Couch, W. J., et al. 2000, ApJ, 542, 673

Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23

Feng, Y., Chu, M.-Y., & Seljak, U. 2016, ArXiv e-prints, arXiv:1603.00476

Filippi, S., Barnes, C., & Stumpf, M. 2011, arXiv preprint arXiv:1106.6280

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

Fry, J. N., & Gaztanaga, E. 1993, ApJ, 413, 447

Gabor, J. M., & Bournaud, F. 2014, MNRAS, 441, 1615

Gabor, J. M., & Davé, R. 2012, MNRAS, 427, 1816

—. 2015, MNRAS, 447, 374

Gabor, J. M., Davé, R., Finlator, K., & Oppenheimer, B. D. 2010, MNRAS, 407, 749

Gabor, J. M., Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 417, 2676

Gallazzi, A., Bell, E. F., Wolf, C., et al. 2009, ApJ, 690, 1883

Gaztañaga, E., Cabré, A., Castander, F., Crocce, M., & Fosalba, P. 2009, MNRAS, 399, 801

Geha, M., Blanton, M. R., Yan, R., & Tinker, J. L. 2012a, ApJ, 757, 85

—. 2012b, ApJ, 757, 85

Gil-Marín, H., Noreña, J., Verde, L., et al. 2014, ArXiv e-prints, arXiv:1407.5668

Gil-Marín, H., Percival, W. J., Verde, L., et al. 2016a, ArXiv e-prints, arXiv:1606.00439

Gil-Marín, H., Verde, L., Noreña, J., et al. 2015, MNRAS, 452, 1914

Gil-Marín, H., Percival, W. J., Brownstein, J. R., et al. 2016b, MNRAS, 460, 4188

Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., et al. 2016, ArXiv e-prints, arXiv:1607.03143

Gu, M., Conroy, C., & Behroozi, P. 2016, ArXiv e-prints, arXiv:1602.01099

Gunn, J. E., & Gott, III, J. R. 1972, ApJ, 176, 1

Guo, H., Zehavi, I., & Zheng, Z. 2012a, ApJ, 756, 127

—. 2012b, ApJ, 756, 127

Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101

Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37

Hahn, C., Scoccimarro, R., Blanton, M. R., Tinker, J. L., & Rodríguez-Torres, S. 2017, MNRAS, arXiv:1609.01714

Hahn, C., Vakili, M., Walsh, K., et al. 2016, ArXiv e-prints, arXiv:1607.01782

Hahn, C., Blanton, M. R., Moustakas, J., et al. 2015, ApJ, 806, 162

Hamilton, A. J. S. 1997, MNRAS, 289, 285

Hamilton, A. J. S., Rimes, C. D., & Scoccimarro, R. 2006, MNRAS, 371, 1188

Harrison, E. R. 1970, Phys. Rev. D, 1, 2726

Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399

Haywood, M., Lehnert, M. D., Di Matteo, P., et al. 2016, A&A, 589, A66

Hearin, A., Campbell, D., Tollerud, E., et al. 2016a, ArXiv e-prints, arXiv:1606.04106

Hearin, A. P., Zentner, A. R., van den Bosch, F. C., Campbell, D., & Tollerud, E. 2016b, MNRAS, arXiv:1512.03050

Heitmann, K., Higdon, D., White, M., et al. 2009, ApJ, 705, 156

Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D. 2010, ApJ, 715, 104

Heitmann, K., Lukić, Z., Fasel, P., et al. 2008, Computational Science and Discovery, 1, 015003

Hermit, S., Santiago, B. X., Lahav, O., et al. 1996, MNRAS, 283, 709

Hermit, S., Santiago, B. X., Lahav, O., et al. 1996, MNRAS, 283, 709

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Hogg, D. W. 1999, ArXiv Astrophysics e-prints, astro-ph/9905116

Hopkins, A. M., & Beacom, J. F. 2006a, ApJ, 651, 142

—. 2006b, ApJ, 651, 142

Hopkins, P. F., Cox, T. J., Kereš, D., & Hernquist, L. 2008a, ApJS, 175, 390

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008b, ApJS, 175, 356

Hu, W., & Eisenstein, D. J. 1998, ApJ, 498, 497

Hu, W., Eisenstein, D. J., & Tegmark, M. 1998, Physical Review Letters, 80, 5255

Hubble, E. P. 1936a, The Realm of the Nebulae (New Haven: Yale University Press)

—. 1936b, The Realm of the Nebulae (New Haven: Yale University Press)

Hyde, J. B., & Bernardi, M. 2009, MNRAS, 394, 1978

Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55

Iovino, A., Cucciati, O., Scodeggio, M., et al. 2010, A&A, 509, A40

Ishida, E. E. O., Vitenti, S. D. P., Penna-Lima, M., et al. 2015a, Astronomy and Computing, 13, 1

—. 2015b, Astronomy and Computing, 13, 1

Jackson, J. C. 1972, MNRAS, 156, 1P

Jarrett, T. H., Chester, T., Cutri, R., et al. 2000, AJ, 119, 2498

Johnston, R. 2011, A&A Rev., 19, 41

Kaiser, N. 1984, ApJ, 284, L9

Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. 2011, ApJ, 730, 61

Kauffmann, G., & Haehnelt, M. 2000, MNRAS, 311, 576

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

Kitaura, F.-S., Rodríguez-Torres, S., Chuang, C.-H., et al. 2016, MNRAS, 456, 4156

Klypin, A., Yepes, G., Gottlober, S., Prada, F., & Hess, S. 2014, ArXiv e-prints, arXiv:1411.4001

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

Knobel, C., Lilly, S. J., Iovino, A., et al. 2012, ApJ, 753, 121

Knox, L. 1995, Phys. Rev. D, 52, 4307

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Kovač, K., Lilly, S. J., Cucciati, O., et al. 2010, ApJ, 708, 505

Kovač, K., Lilly, S. J., Knobel, C., et al. 2014, MNRAS, 438, 717

Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, ApJS, 111, 73

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

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692

Lauer, T. R., Faber, S. M., Richstone, D., et al. 2007, ApJ, 662, 808

Leach, S. M., Cardoso, J.-F., Baccigalupi, C., et al. 2008, A&A, 491, 597

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012a, ApJ, 744, 159

—. 2012b, ApJ, 744, 159

Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, ApJ, 801, 80

Leja, J., van Dokkum, P., & Franx, M. 2013, ApJ, 766, 33

Lemson, G., & Kauffmann, G. 1999, MNRAS, 302, 111

Lesgourgues, J., Mangano, G., Miele, G., & Pastor, S. 2013, Neutrino Cosmology

Lesgourgues, J., & Pastor, S. 2012, ArXiv e-prints, arXiv:1212.6154

—. 2014, New Journal of Physics, 16, 065002

Li, C., & White, S. D. M. 2009, MNRAS, 398, 2177

Lian, J., Yan, R., Zhang, K., & Kong, X. 2016, ApJ, 832, 29

Licquia, T. C., & Newman, J. A. 2015, ApJ, 806, 96

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119

Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218

Lin, C.-A., & Kilbinger, M. 2015a, A&A, 583, A70

—. 2015b, A&A, 583, A70

Lin, C.-A., Kilbinger, M., & Pires, S. 2016a, ArXiv e-prints, arXiv:1603.06773

—. 2016b, ArXiv e-prints, arXiv:1603.06773

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

Magnelli, B., Elbaz, D., Chary, R. R., et al. 2009, A&A, 496, 57

Makarem, L., Kneib, J.-P., Gillet, D., et al. 2014, A&A, 566, A84

Manera, M., Sheth, R. K., & Scoccimarro, R. 2010, MNRAS, 402, 589

Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428, 1036

Maraston, C., Pforr, J., Henriques, B. M., et al. 2013, MNRAS, 435, 2764

Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2009, ApJ, 701, 1765

Markwardt, C. B. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251

Martig, M., Bournaud, F., Teyssier, R., & Dekel, A. 2009, ApJ, 707, 250

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ, 619, L1

Mendel, J. T., Simard, L., Ellison, S. L., & Patton, D. R. 2013, MNRAS, 429, 2212

Miyatake, H., More, S., Mandelbaum, R., et al. 2015, ApJ, 806, 1

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

Moore, B., Lake, G., & Katz, N. 1998, ApJ, 495, 139

Morales, I., Montero-Dorta, A. D., Azzaro, M., et al. 2012, MNRAS, 419, 1187

More, S., van den Bosch, F. C., & Cacciato, M. 2009a, MNRAS, 392, 917

More, S., van den Bosch, F. C., Cacciato, M., et al. 2009b, MNRAS, 392, 801

—. 2013, MNRAS, 430, 747

Morrissey, P., Schiminovich, D., Barlow, T. A., et al. 2005, ApJ, 619, L7

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, ApJ, 767, 50

Muldrew, S. I., Croton, D. J., Skibba, R. A., et al. 2012, MNRAS, 419, 2670

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18

Navarro, J. F., Hayashi, E., Power, C., et al. 2004, MNRAS, 349, 1039

Nishimichi, T., & Taruya, A. 2011, Phys. Rev. D, 84, 043526

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43

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

Norberg, P., et al. 2002, MNRAS, 332, 827

Oemler, A. 1974, ApJ, 194, 1

Oemler, Jr., A. 1974, ApJ, 194, 1

Oh, S. P., Spergel, D. N., & Hinshaw, G. 1999, ApJ, 510, 551

Oka, A., Saito, S., Nishimichi, T., Taruya, A., & Yamamoto, K. 2014, MNRAS, 439, 2515

Okumura, T., Hand, N., Seljak, U., Vlah, Z., & Desjacques, V. 2015, Phys. Rev. D, 92, 103516

Oliver, S., Frost, M., Farrah, D., et al. 2010, MNRAS, 405, 2279

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

—. 1980b, The large-scale structure of the universe

Peebles, P. J. E., & Yu, J. T. 1970, ApJ, 162, 815

Peng, Y., Maiolino, R., & Cochrane, R. 2015, Nature, 521, 192

Peng, Y.-j., & Maiolino, R. 2014, MNRAS, 443, 3643

Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014a, A&A, 571, A16

—. 2014b, A&A, 571, A16

—. 2015a, ArXiv e-prints, arXiv:1502.01589

—. 2015b, ArXiv e-prints, arXiv:1502.01592

—. 2015c, ArXiv e-prints, arXiv:1502.02114

Pope, A. C., & Szapudi, I. 2008, MNRAS, 389, 766

Popping, G., Behroozi, P. S., & Peeples, M. S. 2015, MNRAS, 449, 477

Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, A&A, 523, A13

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

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., & Feldman, M. W. 1999, Molecular biology and evolution, 16, 1791

Reid, B. A., Samushia, L., White, M., et al. 2012, MNRAS, 426, 2719

Riebe, K., Partl, A. M., Enke, H., et al. 2011, ArXiv e-prints, arXiv:1109.0003

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Rodríguez-Torres, S. A., Prada, F., Chuang, C.-H., et al. 2015, ArXiv e-prints, arXiv:1509.06404

Ross, A. J., Percival, W. J., Sánchez, A. G., et al. 2012, MNRAS, 424, 564

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Sánchez, A. G., Kazin, E. A., Beutler, F., et al. 2013, MNRAS, 433, 1202

Sanchez, A. G., Scoccimarro, R., Crocce, M., et al. 2016, ArXiv e-prints, arXiv:1607.03147

Santiago, B. X., & Strauss, M. A. 1992, ApJ, 387, 9

Santini, P., Maiolino, R., Magnelli, B., et al. 2014, A&A, 562, A30

Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889

Schlegel, D., Abdalla, F., Abraham, T., et al. 2011, ArXiv e-prints, arXiv:1106.1706

Scoccimarro, R. 2000, ApJ, 544, 597

—. 2004, Phys. Rev. D, 70, 083007

—. 2015, Phys. Rev. D, 92, 083532

Scoccimarro, R., Colombi, S., Fry, J. N., et al. 1998, ApJ, 496, 586

Scoccimarro, R., & Sheth, R. K. 2002, MNRAS, 329, 629

Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20

Scodeggio, M., Vergani, D., Cucciati, O., et al. 2009, A&A, 501, 21

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

Sefusatti, E., Crocce, M., Scoccimarro, R., & Couchman, H. M. P. 2016, MNRAS, arXiv:1512.07295

Seljak, U. 2000, MNRAS, 318, 203

Sellentin, E., & Heavens, A. F. 2016, MNRAS, 456, L132

Shao, J., & Tu, D. 1995, The Jackknife and Boostrap

Sheldon, E. S., Johnston, D. E., Scranton, R., et al. 2009, ApJ, 703, 2217

Silk, D., Filippi, S., & Stumpf, M. P. H. 2012, ArXiv e-prints, arXiv:1210.3296

Skibba, R. A., Sheth, R. K., Croton, D. J., et al. 2013, MNRAS, 429, 458

Smethurst, R. J., Lintott, C. J., Simmons, B. D., et al. 2015, MNRAS, 450, 435

Smith, G. P., Treu, T., Ellis, R. S., Moran, S. M., & Dressler, A. 2005, ApJ, 620, 78

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

Somerville, R. S., Lemson, G., Sigad, Y., et al. 2001, MNRAS, 320, 289

Sparre, M., & Springel, V. 2016, ArXiv e-prints, arXiv:1604.08205

Springel, V. 2005, MNRAS, 364, 1105

Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776

Steidel, C. C., Adelberger, K. L., Dickinson, M., et al. 1998, ApJ, 492, 428

Stewart, K. R., Bullock, J. S., Wechsler, R. H., & Maller, A. H. 2009, ApJ, 702, 307

Swanson, M. E. C., Tegmark, M., Hamilton, A. J. S., & Hill, J. C. 2008, MNRAS, 387, 1391

Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1

Taruya, A., Koyama, K., Hiramatsu, T., & Oka, A. 2014, Phys. Rev. D, 89, 043509

Taruya, A., Nishimichi, T., & Bernardeau, F. 2013, Phys. Rev. D, 87, 083509

Taruya, A., Nishimichi, T., & Saito, S. 2010, Phys. Rev. D, 82, 063522

Taylor, E. N., Franx, M., van Dokkum, P. G., et al. 2009, ApJ, 694, 1171

Tinker, J., Wetzel, A., & Conroy, C. 2011, ArXiv e-prints, arXiv:1107.5046

Tinker, J., Wetzel, A., Conroy, C., & Mao, Y.-Y. 2016, ArXiv e-prints, arXiv:1609.03388

Tinker, J. L. 2016, ArXiv e-prints, arXiv:1607.06099

Tinker, J. L., Hahn, C., Mao, Y.-Y., Wetzel, A. R., & Conroy, C. 2017, ArXiv e-prints, arXiv:1702.01121

Tinker, J. L., Leauthaud, A., Bundy, K., et al. 2013, ApJ, 778, 93

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005a, ApJ, 631, 41

—. 2005b, ApJ, 631, 41

Tinker, J. L., & Wetzel, A. R. 2010, ApJ, 719, 88

Tinker, J. L., Sheldon, E. S., Wechsler, R. H., et al. 2012, ApJ, 745, 16

Vakili, M., & Hahn, C. H. 2016, ArXiv e-prints, arXiv:1610.01991

Vale, A., & Ostriker, J. P. 2006, MNRAS, 371, 1173

van de Voort, F., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, MNRAS, 415, 2782

van den Bosch, F. C., Mo, H. J., & Yang, X. 2003, MNRAS, 345, 923

van den Bosch, F. C., More, S., Cacciato, M., Mo, H., & Yang, X. 2013, MNRAS, 430, 725

van den Bosch, F. C., Yang, X., Mo, H. J., et al. 2007, MNRAS, 376, 841

Verde, L., Heavens, A. F., Matarrese, S., & Moscardini, L. 1998, MNRAS, 300, 747

Villaescusa-Navarro, F. 2015, ArXiv e-prints, arXiv:1501.04546

Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, Phys. Rev. D, 70, 083511

Weinmann, S. M., van den Bosch, F. C., Yang, X., et al. 2006, MNRAS, 372, 1161

West, A. A. 2005, PhD thesis, University of Washington, Washington, USA

West, A. A., Garcia-Appadoo, D. A., Dalcanton, J. J., et al. 2010, ApJ, 139, 315

Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, MNRAS, 424, 232

Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, MNRAS, 432, 336

—. 2014, MNRAS, 439, 2687

Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013a, ApJ, 764, 116

—. 2013b, ApJ, 764, 116

Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJ, 754, L29

White, M. 2002, ApJS, 143, 241

White, M., Cohn, J. D., & Smit, R. 2010, MNRAS, 408, 1818

White, M., & Scott, D. 1996, Comments on Astrophysics, 18, astro-ph/9601170

White, M., Tinker, J. L., & McBride, C. K. 2014, MNRAS, 437, 2594

Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879

Wilman, D. J., Zibetti, S., & Budavári, T. 2010, MNRAS, 406, 1701

Woo, J., Carollo, C. M., Faber, S. M., Dekel, A., & Tacchella, S. 2016, ArXiv e-prints, arXiv:1607.06091

Wyder, T. K., Martin, D. C., Schiminovich, D., et al. 2007, ApJS, 173, 293

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

—. 2008, ApJ, 676, 248

—. 2009, ApJ, 693, 830

Yesuf, H. M., Faber, S. M., Trump, J. R., et al. 2014, ApJ, 792, 84

Yoon, J. H., Schawinski, K., Sheen, Y.-K., Ree, C. H., & Yi, S. K. 2008, ApJS, 176, 414

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172

Zehavi, I., et al. 2002, ApJ, 571, 172

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1

—. 2011, ApJ, 736, 59

Zeldovich, Y. B. 1972, MNRAS, 160, 1P

Zentner, A. R., Hearin, A., van den Bosch, F. C., Lange, J. U., & Villarreal, A. 2016, ArXiv e-prints, arXiv:1606.07817

Zhai, Z., Tinker, J. L., Hahn, C., et al. 2016, ArXiv e-prints, arXiv:1607.05383

Zhao, G.-B., Saito, S., Percival, W. J., et al. 2013, MNRAS, 436, 2038

Zheng, Z., Coil, A. L., & Zehavi, I. 2007a, ApJ, 667, 760

—. 2007b, ApJ, 667, 760

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005a, ApJ, 633, 791

—. 2005b, ApJ, 633, 791

Zu, Y., & Mandelbaum, R. 2015, MNRAS, 454, 1161