Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 3 October 2017 (MN LaTeX style file v2.2)

Forward Modeling Galaxy Size in a Cosmological Context

Andrew Hearin[1], Peter Behroozi[2], Andrey Kravtsov[3], Benjamin Moster[4]
[1] *Argonne National Laboratory, Argonne, IL, USA 60439, USA*
[2] *Department of Physics, University of Arizona, 1118 E 4th St, Tucson, AZ 85721 USA*
[3] *Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637 USA*
[4] *Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 München, Germany*

3 October 2017

ABSTRACT

We derive empirical modeling constraints on the connection between dark matter halos and the half-light radius of galaxies, $R_{1/2}$. Novel to this work, we study galaxy size using new SDSS measurements of the $R_{1/2}-$dependence of galaxy clustering. Smaller galaxies cluster stronger relative to larger galaxies of the same stellar mass, a new result. We use `Halotools` to test a collection of forward models of galaxy size to identify the qualitative ingredients needed to reproduce the observed clustering. We show that the $R_{1/2}-$dependence of galaxy clustering is largely driven by centrals being larger than satellite galaxies of the same halo mass. Models in which $R_{1/2}$ is determined by stellar mass M_* exhibit qualitatively discrepant clustering properties from SDSS galaxies. Models where $R_{1/2}$ is linearly proportional to halo virial radius $R_{\rm vir}$ at the time of peak halo mass are much more successful, provided that scatter in $R_{1/2}$ is significantly correlated with the halo scale radius $R_{\rm s}$ at that time. Together with the result that stellar mass stripping of satellites has only a mild impact on $R_{1/2}-$dependent clustering, this suggests that the relative size of centrals and satellites is already in place at the time of satellite infall, and supports the notion that L_* galaxy and halo profiles co-evolve across many Gyr of cosmic time.

1 INTRODUCTION

Some introduction goes here.

2 DATA AND SIMULATIONS

Our galaxy sample comes from the catalog of SDSS galaxy profile decompositions provided by Meert et al. (2015). This catalog is based on Data Release 10 of the Sloan Digital Sky Survey (SDSS, Ahn et al. 2014), with improvements to the photometry pipeline and light profile fitting methods (Vikram et al. 2010; Bernardi et al. 2013, 2014; Meert et al. 2013). In the version of this catalog that we use, two-dimensional $r-$band profiles were fit with a two-component de Vaucouleurs + exponential profile to determine the half-light radius $R_{1/2}$. We apply the Bell et al. (2003) mass-to-light ratio to the $r-$band flux and $g-r$ colors in this catalog to obtain an estimate for the total stellar mass M_* of every galaxy. We calculate two-point clustering $w_{\rm p}$ of our SDSS galaxy sample using line-of-sight projection of $\pi_{\rm max} = 20$Mpc using the `correl` program in `UniverseMachine`.

As the bedrock of our modeling, we use the catalog of `Rockstar` subhalos identified at $z = 0$ in the Bolshoi-Planck simulation (Klypin et al. 2011; Behroozi et al. 2013,?; Riebe et al. 2013; Rodríguez-Puebla et al. 2016). The particular version of the catalog we use is made publicly available through `Halotools` (Hearin et al. 2016), with `version_name` = 'halotools_v0p4'.

We additionally explore the potential existence of satellite galaxies that reside in subhalos that are not identified by halo-finder to the present day, so-called "orphan galaxies" (see, e.g., Campbell et al. 2017). We use an extension of `Consistent Trees` that models the evolution of subhalos after disruption. The phase space evolution of orphans is approximated by following a point mass evolving in the host halo potential according to the orbital parameters of the subhalo at the time of disruption; the evolution of subhalo mass and circular velocity is approximated using the semi-analytic model presented in Jiang & van den Bosch (2014).

For mock galaxies, to compute galaxy clustering we employ the distant observer approximation by treating the simulation $z-$axis as the line-of-sight. We compute $w_{\rm p}$ using the `mock_observables.wp` function in `Halotools`, which is a python implementation of the algorithm in the `Corrfunc` C library (Sinha & Garrison 2017).

All numerical values of $R_{1/2}$ will be quoted in physical kpc, and all values of M_* and $M_{\rm halo}$ in $M_{\odot}$, assuming $H_0 = 67.8$ km/s $\equiv 100h$ km/s, the best-fit value from

Planck Collaboration et al. (2016). To scale stellar masses to "$h = 1$ units" (Croton 2013), our numerically quoted values for M_* should be multiplied by a factor of h^2, while our halo masses and distances should be multiplied by a factor of h.

3 GALAXY-HALO MODEL

3.1 Stellar mass model

We map M_* onto subhalos with the best-fit stellar-to-halo mass relation from Moster et al. (2013):

$$\langle M_*/M_{\rm halo}\rangle = 2N\left[(M_{\rm halo}/M_1)^{-\beta} + (M_{\rm halo}/M_1)^{\gamma}\right]^{-1}. \quad (1)$$

For halo mass $M_{\rm halo}$ we use $M_{\rm peak}$, the largest value of $M_{\rm vir}$ ever attained along the main progenitor branch of the subhalo.

The values of the best-fit parameters in Moster et al. (2013) were fit to a stellar mass function (SMF) with values $M_*^{\rm MPA-JHU}$ based on the MPA-JHU catalog (Kauffmann et al. 2003; Brinchmann et al. 2004), which differs from the SMF in our galaxy sample (see, e.g., Bernardi et al. 2014). We account for this difference by manually tabulating the median value $\langle M_*^{\rm Meert+15}|M_*^{\rm MPA-JHU}\rangle$ in logarithmic bins spanning $9 < \log_{10} M_*^{\rm MPA-JHU}/M_{\odot} < 12$, and applying the median correction to the Monte Carlo realization of the mock galaxy sample. This results in a typical boost of ~ 0.25 dex at $M_*^{\rm MPA-JHU} \approx 10^{9.75}M_{\odot}$, and ~ 0.4 dex at $M_*^{\rm MPA-JHU} \approx 10^{11.5}M_{\odot}$.

3.2 Galaxy size models

In §4, we calculate predictions for the $R_{1/2}-$dependence of galaxy clustering for several different kinds of empirical models, described in turn below.

3.2.1 M_*-only model

In the first class of models we explore, we suppose that stellar mass M_* is the statistical regulator of $R_{1/2}$, so that galaxy sizes are drawn from a log-normal distribution centered at $\langle R_{1/2}|M_*\rangle$, where M_* derives from the Moster et al. (2013) relation described above in §3.1. To implement this model, for simplicity we directly tabulate $\langle R_{1/2}|M_*\rangle$ from the data, rather than pursue a parametric form (see, e.g., Zhang & Yang 2017).

3.2.2 $R_{\rm vir}-$only model

In Kravtsov (2013), it was found that if a stellar-to-halo mass relation is inverted to map halo mass estimates $M_{\rm halo}$ onto SDSS galaxies, and then the $M_{\rm halo} - R_{\rm vir}$ relation is applied to map values of $R_{\rm vir}$ onto the galaxies, then the resulting $R_{1/2} - R_{\rm vir}$ relation of SDSS galaxies exhibits the following linear scaling across a wide range of stellar mass:

$$R_{1/2} = 0.0125R_{\rm vir} \quad (2)$$

Motivated by the simplicity of this scaling relation, we transform the Kravtsov (2013) into a forward model using `Halotools`. For the virial radius of halos and subhalos, we use $R_{\rm M_{peak}}$, the value of $R_{\rm vir}$ in physical units of kpc measured at the time of peak subhalo mass, defined by

$$M_{\rm peak} \equiv \frac{4\pi}{3}R^3_{\rm M_{peak}}\Delta_{\rm vir}(z_{M_{\rm peak}})\rho_{\rm m}(z_{M_{\rm peak}}), \quad (3)$$

where for $\Delta_{\rm vir}(z_{M_{\rm peak}})$ we use the fitting function to the "virial" definition used in Bryan & Norman (1998). For the model we refer to as the "$R_{\rm vir}-$only model", we add uncorrelated log-normal scatter of $\sigma_{\rm R_{1/2}} = 0.2$ dex to generate a Monto Carlo realization of the model population.

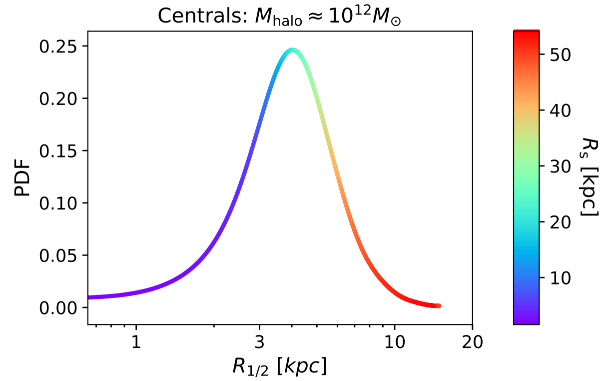

Figure 1. Profile co-evolution model. In this model, the log-normal distribution of galaxy sizes is determined by $R_{\rm M_{peak}}$, the physical size of the halo virial radius at the time the halo reaches its peak mass. Thus log-normal the shape of this curve is determined by $\langle R_{1/2}|R_{\rm M_{peak}}\rangle = 0.0125R_{\rm M_{peak}}$ with 0.2 dex of scatter. The profile co-evolution model further supposes that galaxy and halo profiles co-evolve across cosmic time. Thus within this log-normal distribution, for two halos with the same $R_{\rm M_{peak}}$, halos with larger scale radius at $z_{M_{\rm peak}}$ will host galaxies with above-average $R_{1/2}$, and conversely for halos with smaller values of $R_{\rm s, M_{peak}}$.

3.2.3 Profile co-evolution model

The "profile co-evolution" model is identical to the $R_{\rm vir}-$only model, but the scatter in $R_{1/2}$ at fixed $R_{\rm M_{peak}}$ is no longer purely stochastic, and is instead correlated with $V_{\rm peak}$, defined as the value of $V_{\rm max}$ at $z_{M_{\rm peak}}$. In this way, $R_{\rm M_{peak}}$ statistically determines the distribution of available sizes, but halos with extended dark matter profiles and large scale radii $R_{\rm s}$ host galaxies with above-average sizes, while halos with small $R_{\rm s}$ host smaller galaxies. See Figure 1 for a visual illustration of this residual correlation. We implement scatter correlations using the `halotools.empirical_models.conditional_abunmatch` function, which generalizes the Conditional Abundance Matching technique described in Hearin et al. (2014).

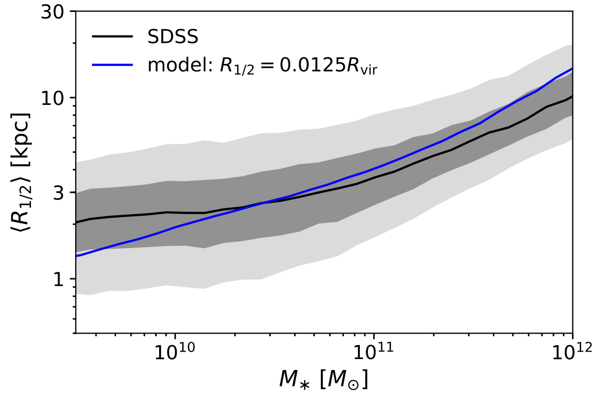

Figure 2. The black curve shows the median $R_{1/2} - M_*$ relation SDSS galaxies as measured in Meert et al. (2015). The two gray bands enveloping the black curve show the 50% and 90% percentile regions. The blue curve shows profile co-evolution model in which $\langle R_{1/2} | R_{\rm vir} \rangle = 0.0125 R_{\rm vir}$, as in Kravtsov (2013). This figure confirms that a linear relationship between $R_{\rm vir}$ and $R_{1/2}$, convolved against the nonlinear relationships between $R_{\rm vir}, M_{\rm halo}$ and M_*, predicts the characteristic curvature in the relation $\langle R_{1/2} | M_* \rangle$ over a wide range in mass.

3.2.4 M_*-stripping model

As we will show in §4, the chief ingredient needed to recover the observed clustering properties of galaxies is that satellites need to be smaller than centrals of comparable halo mass. Thus it is natural to consider a class of models in which stellar mass is stripped from satellite galaxies after infall.

The basis of this class of models is the fitting function presented in Smith et al. (2016), which was calibrated by studying stellar mass loss in a suite of high-resolution hydrodynamical simulations. In this model, f_* quantifies the fraction of stellar mass lost as a function of $f_{\rm DM}$, the amount of dark matter that has been stripped since infall:

$$f_* = 1 - \exp(-14.2 f_{\rm DM}) \tag{4}$$

For $f_{\rm DM}$ we use the ratio of present-day subhalo mass divided by the peak mass, $M_{\rm vir}/M_{\rm peak}$. If we denote the post-stripping stellar mass as M_*', then we have $M_*' \equiv f_* M_*$, where M_* is given by Eq. 1. We then calculate the post-stripping radius by interpolating $\langle R_{1/2}' | M_*' \rangle$ directly from SDSS data.

4 RESULTS

In Figure 2 we show the scaling of galaxy size $R_{1/2}$ with M_*. The black curve enveloped by the gray bands show the scaling relation for our SDSS galaxy sample, while the blue curve shows the median relation $\langle R_{1/2} | M_* \rangle$ implied by the profile co-evolution model described in §3. This figure shows that models in which $R_{1/2} \propto R_{\rm vir}$ can naturally give rise to the characteristic curvature in the $\langle R_{1/2} | M_* \rangle$ relation, confirming the results in Kravtsov (2013) in a forward modeling context.

In Figure 3 we present new measurements of the $R_{1/2}-$dependence of projected galaxy clustering, $w_{\rm p}(r_{\rm p})$.

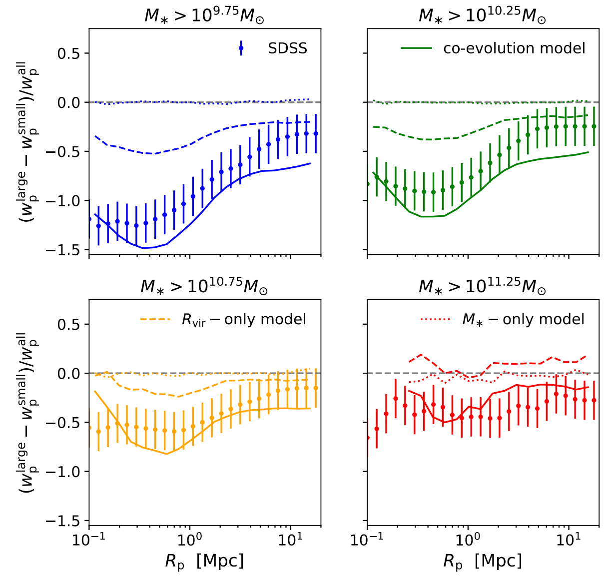

Figure 3. $R_{1/2}-$**dependence of galaxy clustering.** Points with error bars show new SDSS measurements of the $R_{1/2}-$dependence of projected galaxy clustering, $w_{\rm p}$. We define a galaxy as "large" or "small" according to whether it is above or below the median size for its stellar mass, so that in each panel, the stellar mass functions of the "large" and "small" subsamples are identical, as described in the text. The y-axis shows *clustering strength ratios,* so that, for example, a y-axis value of -0.5 corresponds to small galaxies being 50% more strongly clustered than large galaxies of comparable stellar mass. Each panel shows results separately for different volume-limited M_*-threshold samples; predictions of three different models are shown in each panel. See §3.2 for a description of each model.

Because galaxy clustering has well-known dependence upon M_* that is not the subject of this work, we wish to remove this influence and focus purely on the relationship between $R_{1/2}$ and $w_{\rm p}(r_{\rm p})$. To do so, we determine the value $\langle R_{1/2} | M_* \rangle$ by computing a sliding median of $R_{1/2}$, calculated using a window of width $N_{\rm gal} = 1000$. Each galaxy is categorized as either "large" or "small" according to whether it is above or below the median value appropriate for its stellar mass. Using this technique, we stress that for any M_*-threshold sample, the SMF of the "large" and "small" subsamples are identical, by construction.

We measure $w_{\rm p}(r_{\rm p})$ separately for large and small subsamples for four different M_* thresholds, $M_* > 10^{9.75} M_\odot$, $M_* > 10^{10.25} M_\odot$, $M_* > 10^{10.75} M_\odot$, and $M_* > 10^{11.25} M_\odot$. We make the same measurements for each volume-limited M_*-threshold sample *without* splitting on size, giving us measurements $w_{\rm p}^{\rm all}, w_{\rm p}^{\rm large}$, and $w_{\rm p}^{\rm small}$ for each threshold sample. This allow us to compute the ratio $(w_{\rm p}^{\rm large} - w_{\rm p}^{\rm small})/w_{\rm p}^{\rm all}$, which we refer to as *the $R_{1/2}$ clustering ratio.* These ratios are the measurements appearing on the y-axis in each panel of Figure 3. Points with jackknife-estimated error bars show SDSS measurements, solid curves show the clustering ratios of model galaxies as predicted by the models described in §3.

The salient feature of the clustering ratio measurements is that they are negative: small galaxies cluster more strongly than large galaxies of the same stellar mass, a new result. This feature also holds true for galaxies predicted by the $R_{\rm vir}-$only model. This result may be surprising, since $R_{1/2} \propto R_{\rm vir}$, halo mass $R_{\rm vir} \propto M_{\rm halo}^{1/3}$, and clustering strength increases with $M_{\rm vir}$. Based on this simple argument, one would expect the opposite trend to the measurements shown here.

A straightforward resolution to this puzzle is shown in Figure 4, which compares the $R_{1/2}$ distributions of central, satellite, and splashback galaxies with the same halo mass $M_{\rm peak} \approx 10^{12} M_{\odot}$. A "splashback" galaxy is defined as a present-day central that used to be a satellite, i.e., its main progenitor halo passed inside the virial radius of a larger halo at some point in its past history. On the other hand, we define a "true central" as a galaxy that has never been a satellite.

In the $R_{\rm vir}-$only model, satellite and splashback galaxies are smaller than centrals of the same halo mass due to the physical size of their halo being smaller at earlier times $z_{M_{\rm peak}}$. There are two distinct reasons why this feature results in small galaxies being more strongly clustered relative to larger galaxies of the same mass. First, satellite galaxies statistically occupy higher mass host halos that are more strongly clustered. In models where satellites are smaller than centrals, for any given $M_{*}-$threshold the "small" subsample will naturally have a higher satellite fraction, resulting in a negative clustering ratio as seen in SDSS data. Second, at fixed mass, halos of L_{*} galaxies that form earlier are more strongly clustered, a phenomenon commonly known as *halo assembly bias*. Since splashback halos are typically earlier forming than true centrals, then models where splashback halos host smaller-than-average galaxies will naturally predict negative clustering ratios.

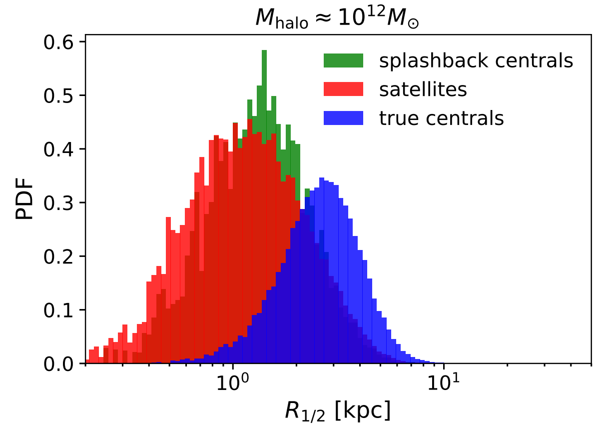

Figure 4. Relative sizes of centrals and satellites. In a narrow bin of halo mass $M_{\rm halo} = M_{\rm peak} \approx 10^{12} M_{\odot}$, we show the distribution of model galaxy sizes for different subpopulations galaxies, as predicted by the $R_{\rm vir}-$only model. The red histogram shows the sizes of satellites; the blue histogram shows host halos that have never passed inside the virial radius of a larger halo ("true centrals"); the green histogram host halos that were subhalos inside a larger at some point in their past history ("splashback halos"). In the $R_{\rm vir}-$only model, galaxy size is set by the *physical* size of the virial radius at the time the halo attains its peak mass, naturally resulting in smaller sizes for satellites and backsplash centrals relative to true centrals of the same $M_{\rm peak}$. The central/satellite size differences shown here are not enough to correctly predict the observed clustering ratios (see Figure 3), motivating the $M_{*}-$stripping and profile co-evolution models.

5 DISCUSSION

5.1 Progression from Backwards to Forwards Modeling

5.2 Implications for Satellite Mass Loss

5.3 Future Directions for Empirical Modeling of Morphology

6 CONCLUSIONS

6.1 Summary

ACKNOWLEDGMENTS

APH thanks John Baker for the *Toejam & Earl* soundtrack.

REFERENCES

Ahn C. P., Alexandroff R., Allende Prieto C., Anders F., Anderson S. F., Anderton T., Andrews B. H., Aubourg É., Bailey S., Bastien F. A., et al. 2014, ApJS , 211, 17
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ , 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013, ApJ , 763, 18
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS , 149, 289
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS , 436, 697
Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS , 443, 874
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS , 351, 1151
Bryan G. L., Norman M. L., 1998, ApJ , 495, 80
Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villarreal A., 2017, ArXiv:1705.06347
Croton D. J., 2013, PASA, 30, e052
Hearin A., Campbell D., Tollerud E., et al., 2016, ArXiv e-prints
Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, MNRAS , 444, 729
Jiang F., van den Bosch F. C., 2014, ArXiv e-prints
Kauffmann G., Heckman T. M., White S. D. M., et al., 2003, MNRAS , 341, 33
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ , 740, 102
Kravtsov A. V., 2013, ApJL , 764, L31
Meert A., Vikram V., Bernardi M., 2013, MNRAS , 433, 1344

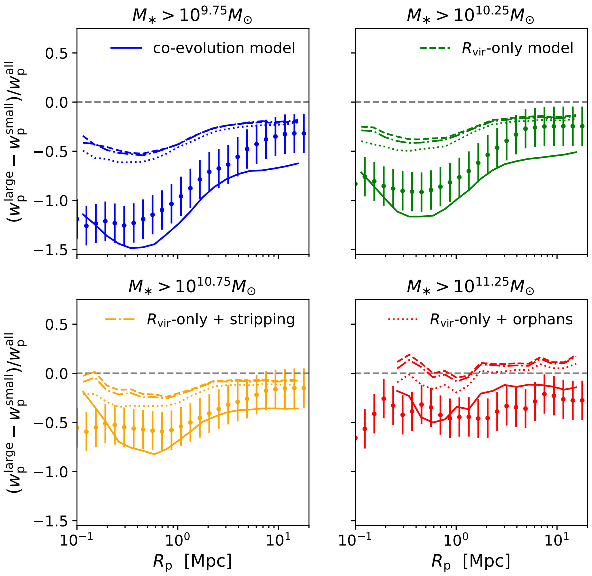

Figure 5. Impact of mass loss and orphan satellites. Here we compare our fiducial model, . . .

Meert A., Vikram V., Bernardi M., 2015, MNRAS , 446, 3943

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

Planck Collaboration Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., et al. 2016, AAP, 594, A13

Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottlöber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2013, Astronomische Nachrichten, 334, 691

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS , 462, 893

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

Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, ApJ , 833, 109

Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS , 409, 1379

Zhang Y., Yang X., 2017, ArXiv e-prints