Creating Mock Galaxy Catalogues From Dark Matter Simulations

Mladen Ivkovic

6. September 2018

Mock Galaxy Catalogues

Mock galaxy catalogues, artificial galaxy catalogues created using simulations, have many uses in astronomy and cosmology:

- Direct comparison to observations to test theories and assumptions
- Estimate statistical and systematic errors
- Planning future surveys, what can we expect to find?
- Development of analysis tools for observational data

Creating Mock Galaxy Catalogues

There are various approaches on how to obtain mock galaxy catalogues.

Essentially, one can differentiate between two basic approaches:

- Physical models
- Empirical models

Physical Models

Simulate or parametrise the physics of galaxy formation (e.g. gas cooling, star formation, feedback processes).

Examples:

- Full hydrodynamical simulation of the Universe
- Semi-Analytic models: approximate some processes with analytical prescriptions, constrain parameters of prescriptions with empirical data

Problems:

- Some processes are still not fully understood
- Computationally expensive

Empirical Models

Don't try to explain how exactly galaxies form, just to predict where and how many galaxies should be and what properties they should have based on some very simple assumptions.

Key assumption: Galaxies form in condensates ("*haloes*") of dark matter.

ΛCDM model states that the Universe is made up from $\sim 5\%$ baryonic ("ordinary") matter, $\sim 25\%$ dark matter and $\sim 70\%$ dark energy. Over time, the matter in the Universe will clump together and galaxies form in centres of these clumps.

DMO Simulations

For computational efficiency, neglect baryonic effects in the Universe and replace the baryonic matter with dark matter in order to preserve the total matter content ("*dark matter only*" simulations).

With DMO simulations available, a galaxy-halo connection is necessary.

Empirical galaxy-halo connections

- HOD: gives probability distribution of the number of galaxies that meet some criteria, e.g. minimal mass threshold
- Conditional stellar mass/luminosity functions: Specify full distribution of galaxy masses/luminosities for a given halo mass
- Abundance Matching: Most massive galaxy lives in most massive halo. Assign galaxies to haloes by rank-order.

The parameters of these techniques are constrained by observational data to reproduce galaxy distributions and properties.

SMHM relation

With these techniques, a stellar-mass-halo-mass-relation can be parametrised and constrained:

$$\log_{10}(M_*(M_h)) = \log_{10}(\epsilon M_1) + f\left(\log_{10}\left(\frac{M_h}{M_1}\right)\right) - f(0) \quad (1)$$

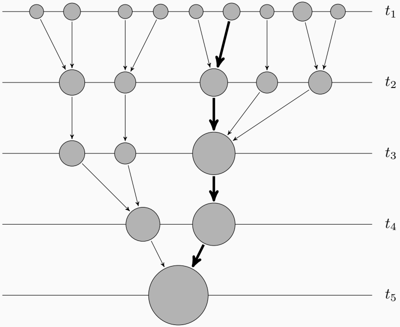
$$f(x) = -\log_{10}(10^{\alpha x} + 1) + \delta \frac{[\log_{10}(1 + \exp(x))]^{\gamma}}{1 + \exp(10^{-x})} \quad (2)$$

(Behroozi, Wechsler, and Conroy 2013)

However, to get accurate galaxy catalogues, some more things need to be considered.

Hierarchical Structure Formation

In the hierarchical structure formation picture, large haloes are thought to form mainly through consecutive merging events of smaller haloes and accretion of matter too small to be recognised as a halo.



A merger tree showing the formation history of some halo over cosmic time through a series of merging events with $t_1 < t_2 < t_3 < t_4 < t_5$.

Due to the hierarchical structure formation, haloes are expected to be spatially nested, i.e. to contain subhaloes.

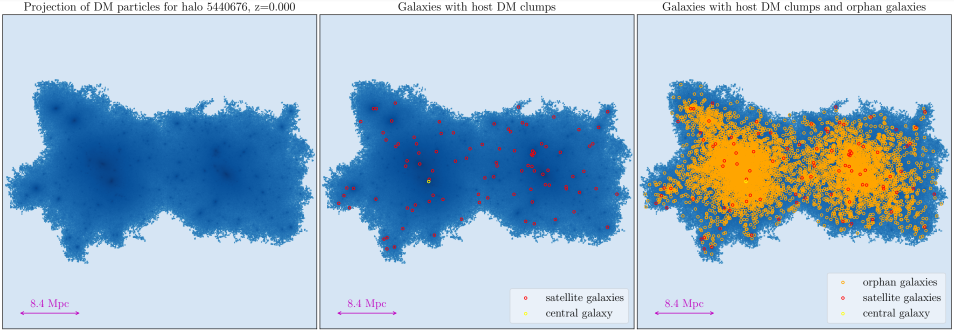
Over time, the material from subhaloes within a host halo will be stripped due to tidal forces. The tidal stripping affects the outer regions of the subhalo much stronger than regions close to its centre, where it's galaxy is located.

$\Rightarrow$ instead of current mass M_h, use peak mass or mass at accretion for subhaloes

What if subhaloes are stripped of so much mass that they can't be identified as substructure in the host's density field any longer?

$\Rightarrow$ Their galaxy still needs to be kept track of (”*orphan galaxies*”)

Orphans and Merger Trees



Projection along the z-axis of the most massive halo from a 512^3 particle simulation, with positions of galaxies at $z = 0$ shown

For accurate mock galaxy catalogues, the formation history of dark matter haloes is necessary.

$\Rightarrow$ need "*merger trees*", which show that formation history

Goals of This Thesis

- Implement an algorithm into RAMSES (Teyssier, R. 2002) to create mock galaxy catalogues from dark matter simulations on the fly and in parallel.
- For mock galaxy catalogues, first dark matter halo merger trees must be obtained, also on the fly and in parallel.

Making Merger Trees

Haloes are defined at distinct snapshots.

Snapshots correspond to particular values of cosmic time and contain the particle IDs, mass, location & velocity for each dark matter particle in the simulation.

The aim of a merger tree code is to link clumps (haloes, subhaloes) from an earlier snapshot to the clumps of the consecutive snapshot, i.e. to find the descendants of the clumps of the earlier snapshot within the consecutive snapshot, thus enabling the tracking of growth and merges of haloes in a simulation.

Making Merger Trees

How to link progenitors and descendants?

$\Rightarrow$ trace particles of clumps by their unique particle ID.

In essence: "in which clumps did particles of a progenitor clump end up in?"

Trace up to a maximal number, n_{mb}, of most tightly bound particles of each clump. The most tightly bound particles are expected to more likely remain in the same clump in the following snapshot.

Making Merger Trees

To obtain a merger *tree*, as opposed to a merger graph, each progenitor may have exactly one direct descendant clump. Descendants however may have multiple direct progenitors. The other direct progenitors of this descendant are then assumed to have merged into the main direct progenitor to form the descendant.

Both progenitors and descendants may have multiple candidates to choose from. To identify best progenitor/descendant candidate: Maximise merit function $\mathcal{M}$

Making Merger Trees: Merit Function

$$\mathcal{M}(A, B) = \frac{n_{A \cap B}}{\frac{m_>}{m_<} - 1} \tag{3}$$

for each progenitor A and descendant B, with $n_{A \cap B}$ = particles both in A and B and $m_>, m_<$ is the bigger or smaller mass between A, B, respectively.

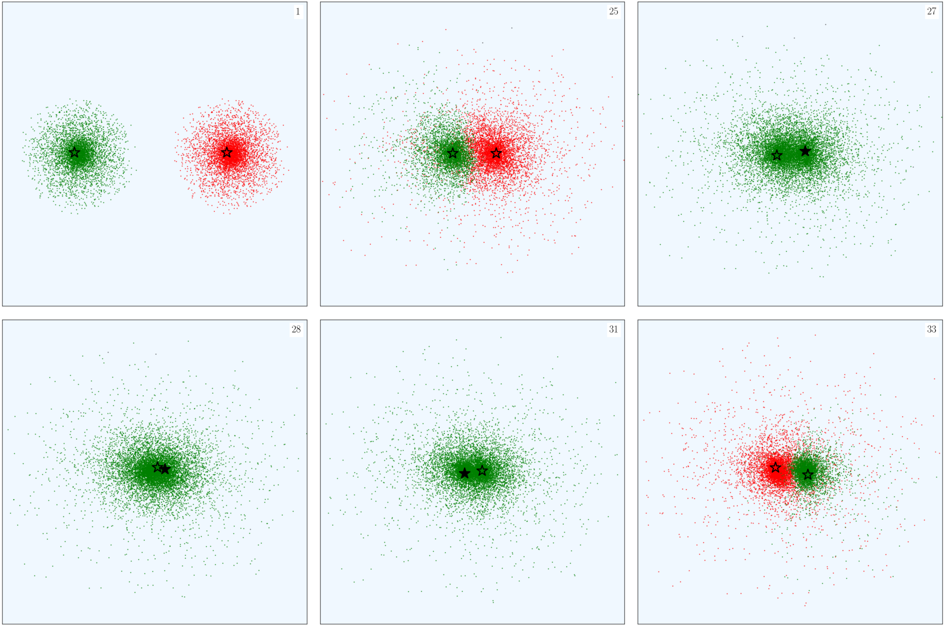
The factor $(m_>/m_< - 1)^{-1}$ was chosen to prefer progenitor-descendant pairs of similar mass to avoid strong mass growth fluctuations.

Making Merger Trees: Jumpers and Orphans

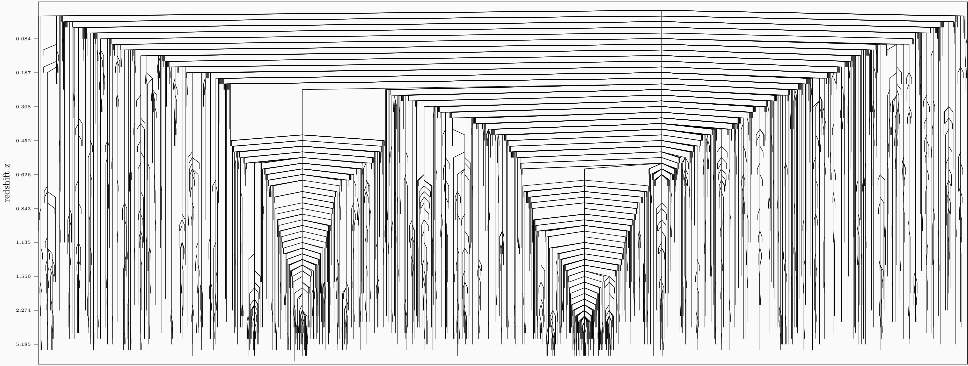
Galaxies will be placed at the position of the most tightly bound particle of each clump once the merger trees have been obtained.

Once a clump is merged into some descendant, the last identifiable “galaxy particle” is tracked as an orphan galaxy in future snapshots.

The orphan galaxy particles are also used to counter failures of the clump finding algorithm: Sometimes substructure is not identified in the density field of a host halo, even though it exists.

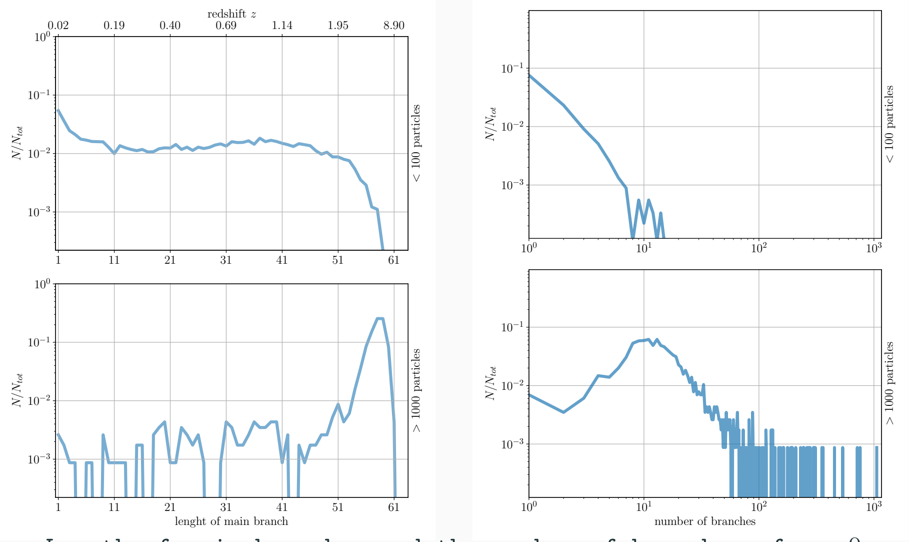

1
25
27
28
31
33

Example of a Resulting Mergertree



The merger tree of the most massive central halo of a $256^3 \approx 1.7 \times 10^7$ particle simulation. The redshift at the time of the snapshot is given on the y-axis, the x-axis has no physical meaning.

Tree Statistics



Length of main branches and the number of branches of $z = 0$ clumps of a 256^3 particle simulation with $m_p \approx 1.6 \times 10^9 M_\odot$.

What Parameters Give Best Merger Trees?

The exact definition of a halo and particularly for a subhalo is not unique throughout literature and may depend on the application.

What definition is best for reliable merger trees?

2 sets of definitions have been tested:

1. does substructure of subhaloes contribute to subhaloes' mass? (`inclusive`) or not? (`exclusive`)
2. Define bound particle: Is it allowed to leave spatial boundary of clump (`no saddle`) or not? (`saddle`)

Logarithmic Mass Growth

$$\frac{\mathrm{d}\log M}{\mathrm{d}\log t} \approx \frac{(t_{k+1}+t_k)(M_{k+1}-M_k)}{(t_{k+1}-t_k)(M_{k+1}+M_k)} \equiv \alpha_M(k, k+1) \tag{4}$$

where k and $k+1$ are a clump and its descendant, with masses M_k and M_{k+1} at times t_k and t_{k+1}, respectively.

To reduce the range of possible values to the interval $(-1, 1)$, define

$$\beta_M = \frac{2}{\pi}\arctan(\alpha_M) \tag{5}$$

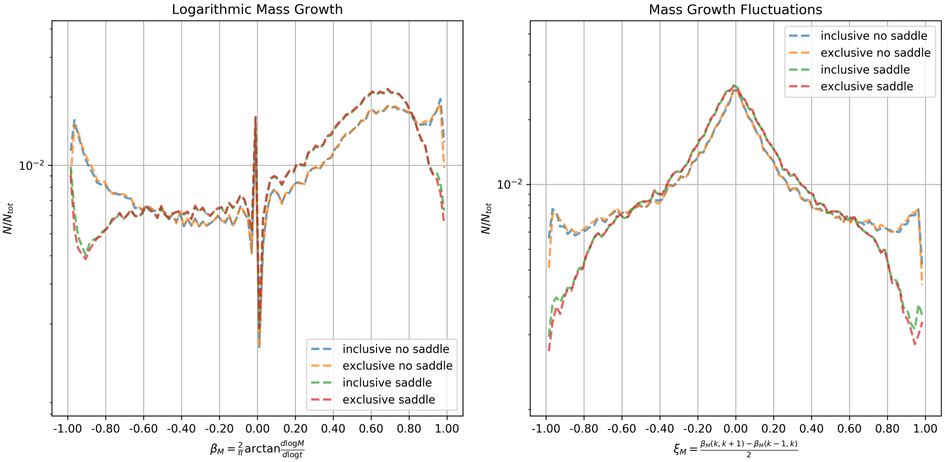
$\beta_M \to \pm 1$ imply $\alpha_M \to \pm\infty$, indicating extreme mass growth or losses.

Mass Growth Fluctuations

$$\xi_M = \frac{\beta_M(k, k+1) - \beta_M(k-1, k)}{2} \tag{6}$$

where $k-1$, k, $k+1$ represent consecutive snapshots. When far from zero, it implies an extreme growth behaviour. For $\xi_M \to \pm 1$, $\beta_M(k, k+1) \to \pm 1$ and $\beta_M(k-1, k) \to \mp 1$, indicating extreme mass loss followed by extreme mass growth for the upper sign, and the opposite behaviour for the lower sign.

Mass Growth and Mass Growth Fluctuations



The distributions are computed as a histogram normalised by the total number of events found throughout the entire simulation. Only clumps with masses above $5 \times 10^{11} M_{\odot}$ were included.

Mock Galaxy Catalogues

With merger trees and a galaxy-halo relation, mock galaxy catalogues can be obtained.

Two sets of mock galaxy catalogues were created and tested. Both contain $512^3 \approx 1.3 \times 10^8$ particles, but different box sizes were used: `G69` spans $69 Mpc$ in each direction, `G100` $100 Mpc$. The mass resolution for particles corresponds to $9.59 \times 10^7 M_\odot$ and $3.09 \times 10^8 M_\odot$, respectively.

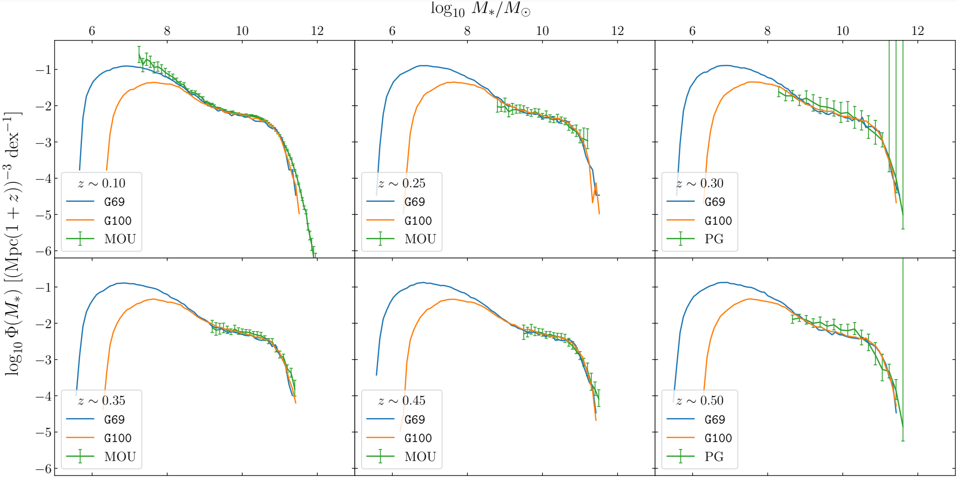
The mass threshold for clumps was chosen to be 10 particle masses.

Testing Mock Galaxy Catalogues

The performed tests of the mock galaxy catalogues are:

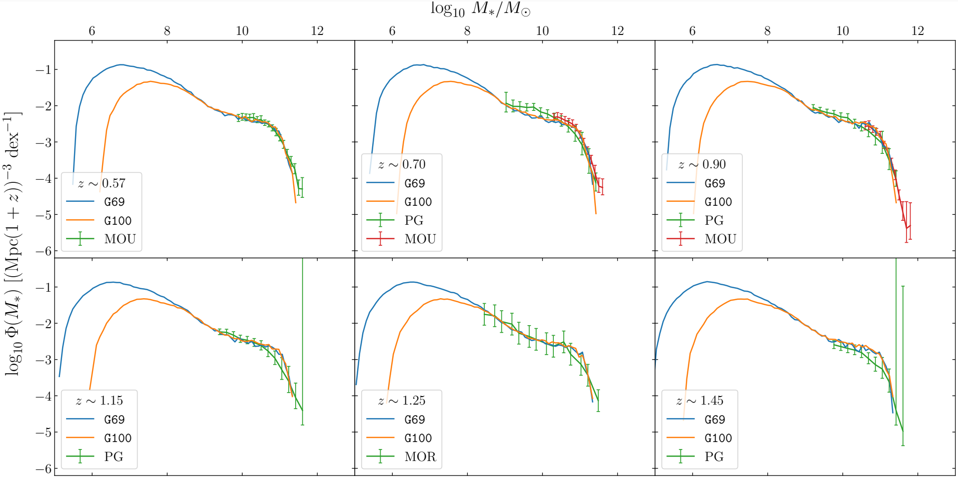
1. Stellar Mass Functions $\Phi(M_*)$
   give number density of central galaxies with stellar mass M_*, directly compared to observational data
2. Clustering Statistics
   The two-point correlation function $\xi(r)$ is commonly used as a measure of galaxy clustering. It can be interpreted as the excess probability to find a galaxy in a volume element at a separation r from another galaxy compared to what is expected for a uniform random distribution.

Stellar Mass Functions: $z = 0 - 0.5$



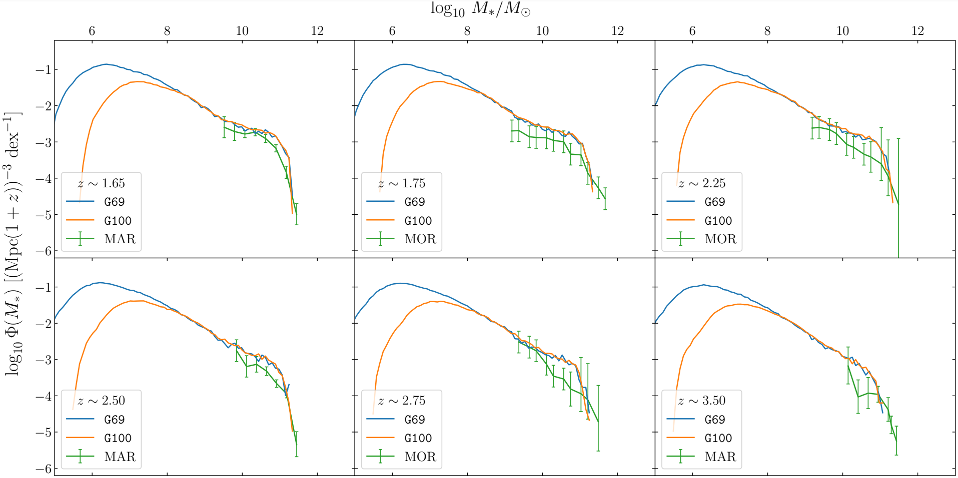
Obtained stellar mass functions $\Phi(M_*)$ of central galaxies for the two simulation datasets G69 and G100, with boxsizes of 69 and 100 Mpc, respectively, compared to observed stellar mass functions.

Stellar Mass Functions: $z = 0.5 - 1.5$



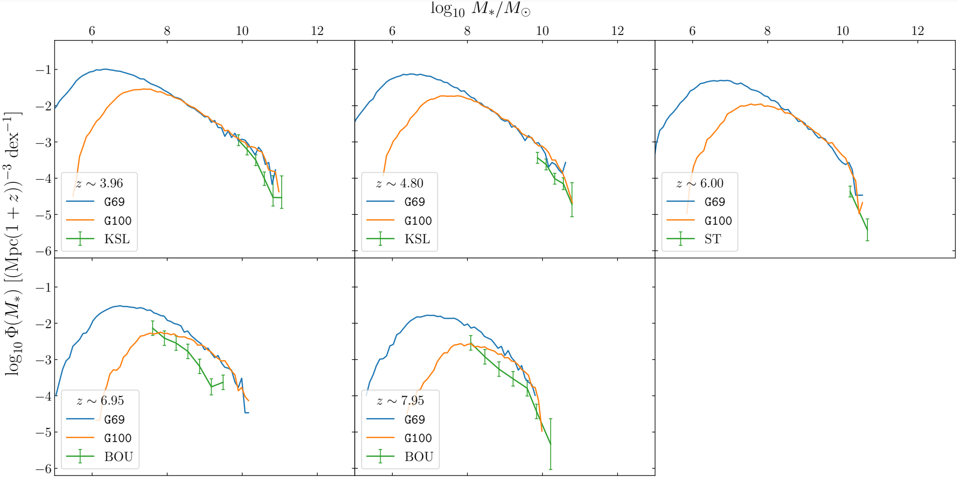
Obtained stellar mass functions $\Phi(M_*)$ of central galaxies for the two simulation datasets G69 and G100, with boxsizes of 69 and 100 Mpc, respectively, compared to observed stellar mass functions.

Stellar Mass Functions: $z = 1.5 - 3.5$



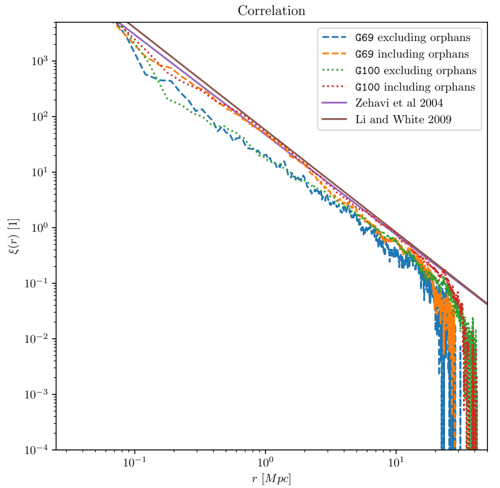
Obtained stellar mass functions $\Phi(M_*)$ of central galaxies for the two simulation datasets G69 and G100, with boxsizes of 69 and 100 Mpc, respectively, compared to observed stellar mass functions.

Stellar Mass Functions: $z = 3.5 - 8$



Obtained stellar mass functions $\Phi(M_*)$ of central galaxies for the two simulation datasets G69 and G100, with boxsizes of 69 and 100 Mpc, respectively, compared to observed stellar mass functions.

Correlation Functions



The obtained 2PCF $\xi(r)$ for the G69 (dashed lines) and G100 (dotted lines) simulations, both including and excluding orphan galaxies, compared to best power law fits of the 2PCF from Li and White 2009 and Zehavi et al. 2004 (solid lines).

Outlook

The results show good agreement with observational data, and should improve with increased resolution.

Currently missing: A merging mechanism for individual orphan galaxies. Options would be:

- Introduce some galaxy-galaxy merging cross section
- Estimate expected merging time for each orphan galaxy individually, e.g. dynamical friction time or fitted merger timescale by Jiang et al. 2008

Sources of SMFs

Redshift interval of observed stellar mass functions that are used for comparison and the abbreviation used in this work as reference.

redshift interval	Reference	Abbreviation
$z \sim 0 - 1$	Moustakas et al. 2013	MOU
$z \sim 0 - 4$	Pérez-González et al. 2008	PG
$z \sim 1 - 3.5$	Mortlock et al. 2011	MOR
$z \sim 1.3 - 4.0$	Marchesini et al. 2009	MAR
$z \sim 4$, $z \sim 5$	Lee et al. 2012	KSL
$z \sim 4 - 6$	Stark et al. 2009	ST
$z \sim 7$, $z \sim 8$	Bouwens et al. 2011	BOU

References I

P. S. Behroozi, R. H. Wechsler, and C. Conroy. "The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8". In: *The Astrophysical Journal* 770, 57 (June 2013), p. 57. DOI: 10.1088/0004-637X/770/1/57. arXiv: 1207.6105 [astro-ph.CO].

R. J. Bouwens et al. "Ultraviolet Luminosity Functions from 132 $z \sim 7$ and $z \sim 8$ Lyman-break Galaxies in the Ultra-deep HUDF09 and Wide-area Early Release Science WFC3/IR Observations". In: *The Astrophysical Journal* 737.2 (2011), p. 90. URL: http://stacks.iop.org/0004-637X/737/i=2/a=90.

C. Y. Jiang et al. "A Fitting Formula for the Merger Timescale of Galaxies in Hierarchical Clustering". In: *The Astrophysical Journal* 675 (Mar. 2008), pp. 1095–1105. DOI: 10.1086/526412. arXiv: 0707.2628.

Kyoung-Soo Lee et al. "How Do Star-forming Galaxies at z > 3 Assemble Their Masses?" In: *The Astrophysical Journal* 752.1 (2012), p. 66. URL: http://stacks.iop.org/0004-637X/752/i=1/a=66.

C. Li and S. D. M. White. "The distribution of stellar mass in the low-redshift Universe". In: *Monthly Notices of the RAS* 398 (Oct. 2009), pp. 2177–2187. DOI: 10.1111/j.1365-2966.2009.15268.x. arXiv: 0901.0706.

D. Marchesini et al. "The Evolution of the Stellar Mass Function of Galaxies from z = 4.0 and the First Comprehensive Analysis of its Uncertainties: Evidence for Mass-Dependent Evolution". In: *The Astrophysical Journal* 701 (Aug. 2009), pp. 1765–1796. DOI: 10.1088/0004-637X/701/2/1765. arXiv: 0811.1773.

References II

Alice Mortlock et al. “A deep probe of the galaxy stellar mass functions at $z \sim 1 - 3$ with the GOODS NICMOS Survey”. In: *Monthly Notices of the Royal Astronomical Society* 413.4 (2011), pp. 2845–2859. DOI: 10.1111/j.1365-2966.2011.18357.x. eprint: /oup/backfile/content_public/journal/mnras/413/4/10.1111/j.1365-2966.2011.18357.x/2/mnras0413-2845.pdf. URL: http://dx.doi.org/10.1111/j.1365-2966.2011.18357.x.

John Moustakas et al. “PRIMUS: Constraints on Star Formation Quenching and Galaxy Merging, and the Evolution of the Stellar Mass Function from z = 0-1”. In: *The Astrophysical Journal* 767.1 (2013), p. 50. URL: http://stacks.iop.org/0004-637X/767/i=1/a=50.

Pablo G. Pérez-González et al. “The Stellar Mass Assembly of Galaxies from z = 0 to z = 4: Analysis of a Sample Selected in the Rest-Frame Near-Infrared with Spitzer”. In: *The Astrophysical Journal* 675.1 (2008), p. 234. URL: http://stacks.iop.org/0004-637X/675/i=1/a=234.

Daniel P. Stark et al. “The Evolutionary History of Lyman Break Galaxies Between Redshift 4 and 6: Observing Successive Generations of Massive Galaxies in Formation”. In: *The Astrophysical Journal* 697.2 (2009), p. 1493. URL: http://stacks.iop.org/0004-637X/697/i=2/a=1493.

Teyssier, R. “Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES”. In: *Astronomy and Astrophysics* 385 (Apr. 2002), pp. 337–364. DOI: 10.1051/0004-6361:20011817. eprint: astro-ph/0111367.

References III

Y. Wang et al. “Sussing merger trees: stability and convergence”. In: *Monthly Notices of the RAS* 459 (June 2016), pp. 1554–1568. DOI: 10.1093/mnras/stw726. arXiv: 1604.01463.

I. Zehavi et al. “On Departures from a Power Law in the Galaxy Correlation Function”. In: *The Astrophysical Journal* 608 (June 2004), pp. 16–24. DOI: 10.1086/386535. eprint: astro-ph/0301280.

Extras

Obtaining 2PCF $\xi(r)$:

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\langle\rho(\mathbf{r})\rangle} - 1 \quad \text{density contrast}$$

$$\delta_{\mathbf{k}} = \frac{1}{V}\int e^{i\mathbf{k}\mathbf{r}}\delta(\mathbf{r})\mathrm{d}^3\mathbf{r} \quad V = L^3 \text{ periodical volume}$$

$$P(k) = V\langle|\delta_{\mathbf{k}}|^2\rangle \quad \text{Power Spectrum}$$

$$\xi(r) = \frac{1}{(2\pi)^3}\int e^{-i\mathbf{k}\mathbf{r}}P(k)\mathrm{d}^3\mathbf{k} \quad \text{2PCF}$$

$$w_p(r_p) = 2\int_0^\infty \mathrm{d}r_{||}\xi\left(\sqrt{r_{||}^2 + r_p^2}\right) \quad \text{Projected Correlation function}$$

$$= 2\int_{r_p}^\infty \mathrm{d}r \frac{r\xi(r)}{\sqrt{r^2 - r_p^2}}$$

Making Merger Trees: Fractures

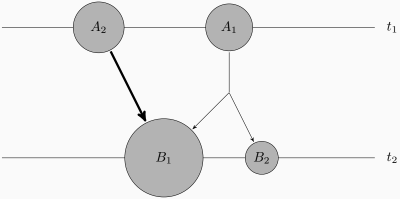
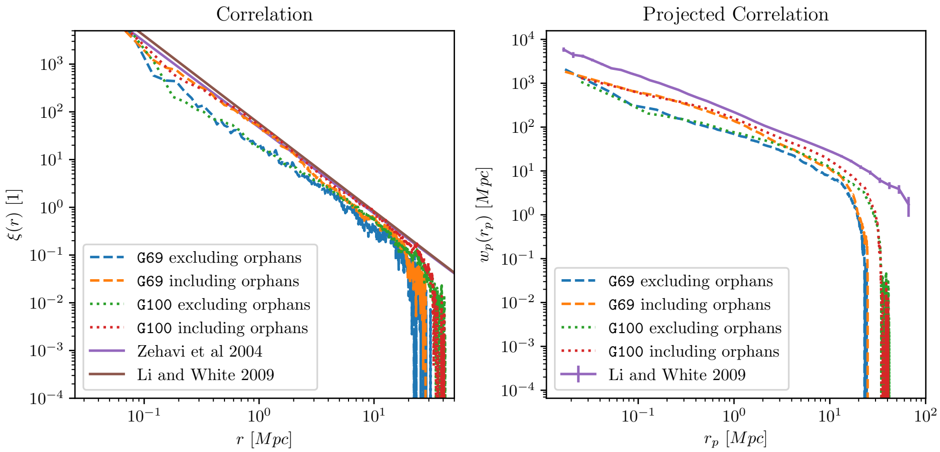


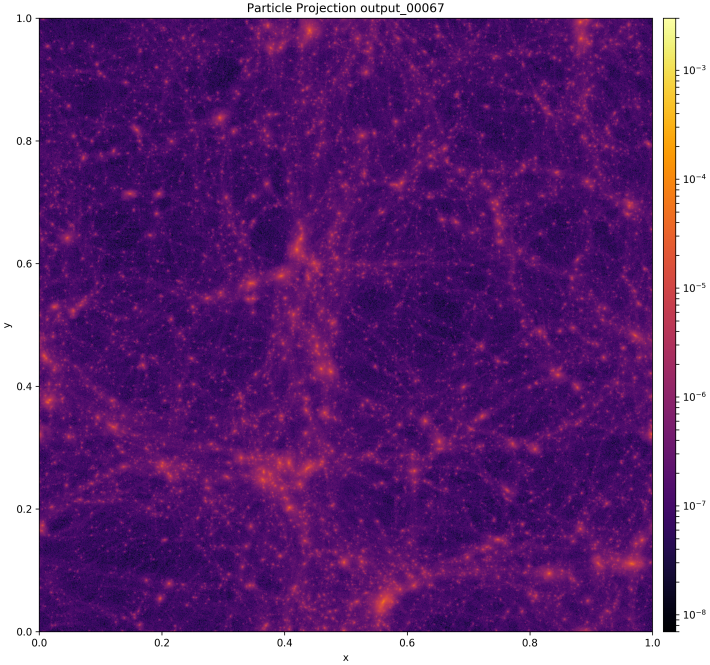
Illustration of a progenitor A_1 at time t_1 which is partially merged into a descendant B_1 at time $t_2 > t_1$, but some other part B_2 isn't. Because A_1 is not the main progenitor of B_1, by assigning its descendant only according to the merit function (3) would not pass on its formation history to B_2, but treat it as newly formed.

$\Rightarrow$ prefer to link progenitors with any descendant candidate instead of merging it into best candidate

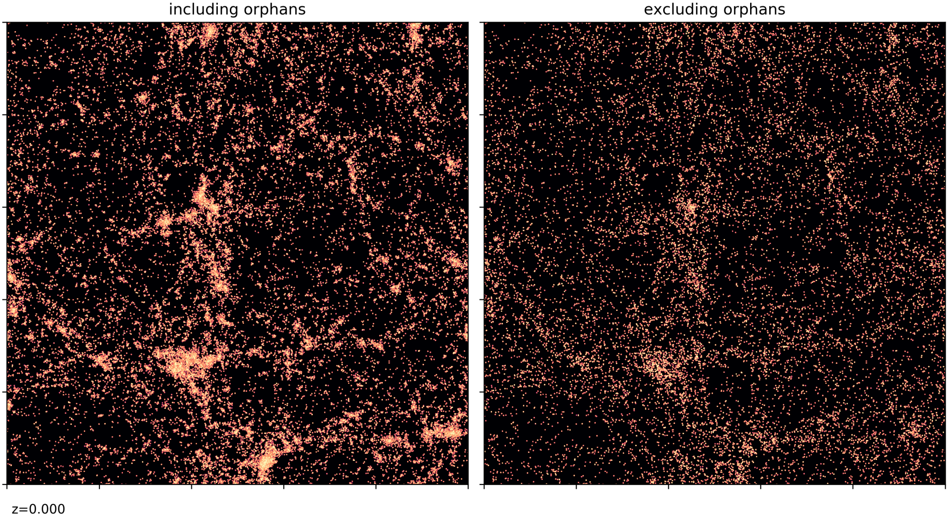
Correlation Functions



The obtained 2PCF $\xi(r)$ and projected correlation function $w_p(r_p)$ for the G69 (dashed lines) and G100 (dotted lines) simulations, both including and excluding orphan galaxies, compared to best power law fits of the 2PCF from Li and White 2009 and Zehavi et al. 2004 and the projected correlation function from Li and White 2009 (solid lines).



Particle Projection output_00067
y
x
0.0
0.2
0.4
0.6
0.8
1.0
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}


including orphans
excluding orphans
z=0.000