

# Gravitational Wave Bias

Nikshay and Taia

July 29, 2025

## Abstract

Your abstract.

## 1 Catalogs

### 1.1 TNG-300 simulation

The IllustrisTNG project provides high-resolution simulation of galaxy formation. We use TNG300-1 simulation which spans the cubic volume of 300 Mpc [1]. Substructure identification is also a part of the IllustrisTNG. They present it in a form of group catalogs, with each containing dark matter halos and galaxies, which are referred to as subhalos. In particular, we analyze group catalogs data on subhalos for snapshots  $z = 0, 0.1, 0.2, 1$ .

### 1.2 Properties of galaxies (subhalos) in the simulation

From the full TNG300 subhalo catalog, we select a subset of fields necessary for our analysis:

- SubhaloFlag – an indication whether subhalos is of cosmological origin. We exclude subhalos with this entity equal to zero from our analysis.
- SubhaloMass – total mass, total stellar mass, and total mass of all black holes. in units of  $10^{10}M_{\odot}/h$
- From SubhaloMassType, which refers to total mass of particles/cells, bound to this subhalo, excluding subhalos of the subhalo, we selected fourth and fifth entries. Fourth entry is the total stellar mass of the subhalo, while fifth entry is the total mass of all black holes within the subhalo.
- SubhaloSFR – total SFR within the subhalo.
- SubhaloStarMetallicity – mass-weighted average metallicity of stars within twice the stellar half mass radius.
- SubhaloGasMetallicity – mass-weighted average metallicity of gas within twice the stellar half mass radius.
- SubhaloVelDisp – the 3D dispersion of velocity of all the member particles/cells divided by  $\sqrt{3}$ .
- SubhaloVmax – peak value of the spherically averaged rotation curve, calculated using all available particle types
- SubhaloSpin – the total 3D angular momentum vector of the subhalo. Calculated as the mass-weighted sum of the cross-products of positions and velocities of each particle/cell within the subhalo.
- SubhaloVel – peculiar velocity of the group. Calculated as the mass-weighted sum of velocities of all particles and cells in this group.
- SubhaloPos – 3D position of the subhalo in the periodic simulation box. Defined by the location of the most gravitationally bound particle (the one with the lowest potential energy). The coordinates are comoving.

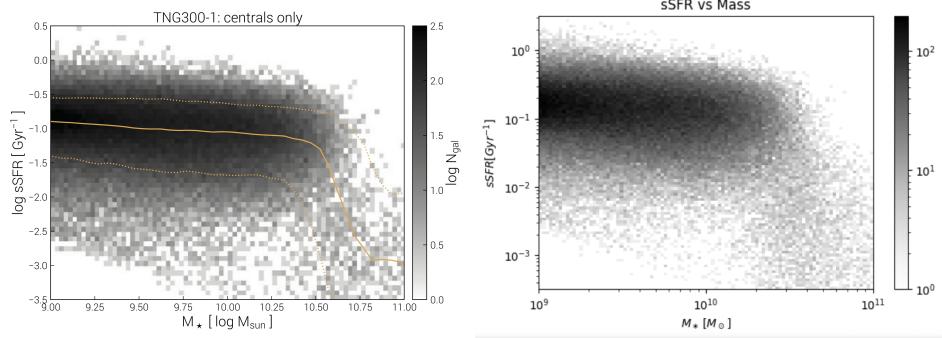


Figure 1: Specific stellar formation rate vs mass. Left - from the literature [1]

- SubhaloGrNr – index that links this subhalo to its parent Friends-of-Friends (FoF) group. A FoF group is a collection of dark matter particles linked by spatial proximity, representing a distinct dark matter halo. Particles of other types are assigned to a FoF group based on the group membership of their nearest dark matter particle.
- SubhaloParent – index pointing to the hierarchical parent of this subhalo. Local to the FoF group. A value of zero indicates no deeper hierarchy, meaning that this subhalo is either central or directly attached to it.

We also saved redshift of the corresponding snapshots. To have a sanity check, we compared our dataset to some literature examples. Comparison of our specific SFR vs stellar mass 2D distribution to the one in [1] is presented in the Figure 1. Comparison of our stellar mass distribution of subhalos with  $M_* > 10^9 M_\odot$  to the one in [2] is presented in the Figure 2. Finally, Figure 3 compares our stellar metallicity vs total mass and stellar mass vs total mass 2D distributions to ones found in [3] and [4].

For further analysis we selected subhalos of stellar mass  $> 10^9 M_\odot$ . Distributions of values of key properties for each snapshot are presented in Figure 4. We binned the star formation rate (SFR), metallicity ( $Z$ ), and stellar mass ( $M_*$ ) into 100 logarithmically spaced interval, spanning the full range for  $Z$  and  $M_*$  and implementing a lower cutoff of  $10^{-3} M_\odot/\text{yr}$  for SFR. The distribution of stellar masses resembles a broken power law. The turning point of this distribution is higher for lower redshifts. Among other visible trends, there are more galaxies with higher SFR for lower redshift, and more galaxies with higher metallicity for lower redshift.

One interesting dependence to investigate is the mean mass dependence on SFR, which we show in Figure 5. We binned SFRs 100 logarithmically spaced intervals, assigned each SFR value to a bin, and computed mean stellar mass within each bin. There is a clear trend of decreasing mass for SFRs below  $\mathcal{O}(1)$  and increasing for mass for higher SFRs. This potentially points at distinction between two groups: quiescent galaxies and star-forming sequence

## 2 Subhaloes power spectrum

We used `CatalogFFTPower` from `pypower` to obtain the power spectrum. To check the procedure, we build an SFR-weighted power spectrum for snapshot at redshift  $z = 1.0$  and compared it to [5]. We converted subhalo positions to cMpc/h and assigned them to a 3D mesh of size  $512^3$  with a total box size of  $205.001 \text{ cMpc}/\text{h}$ . The box size was chosen to just slightly exceed the maximum value of array of coordinates. We use the Triangular Shaped Cloud (TSC) mass assignment scheme with interlacing = 2 to reduce aliasing effects. Then we computed the power spectrum in linear k-bins up to the Nyquist frequency, and we set the shot noise to zero to avoid subtraction for diagnostic purposes. Comparison of our results to [5] are given in the Figure 5.

Further, we unweighted power spectrum for subhalos for  $z = 0.1$  and maximum  $k = 0.3$  with other parameters of `CatalogFFTPower` remaining the same. We calculated total power spectrum, as well as power spectrum for several mass bins. Results are presented in the Figure

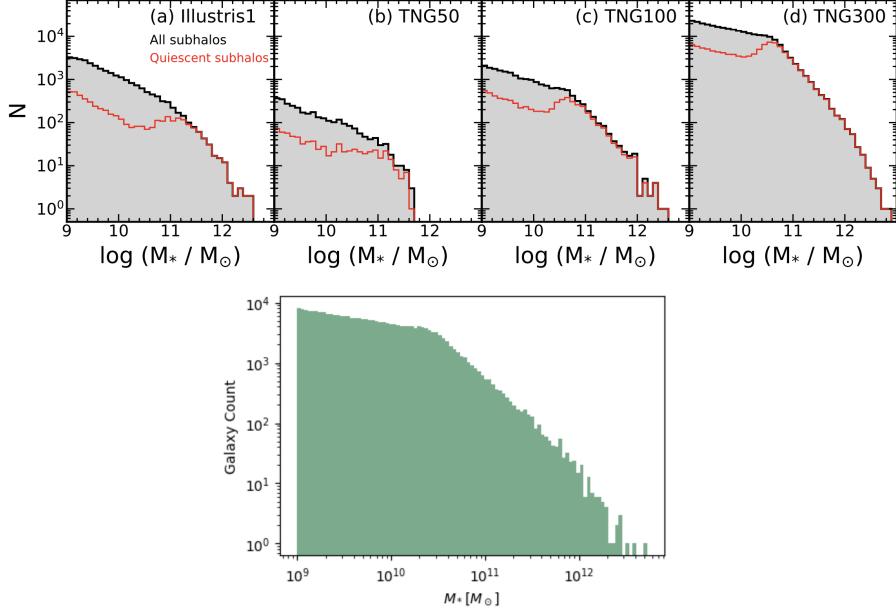


Figure 2: Mass distribution. Upper - from the literature [2]

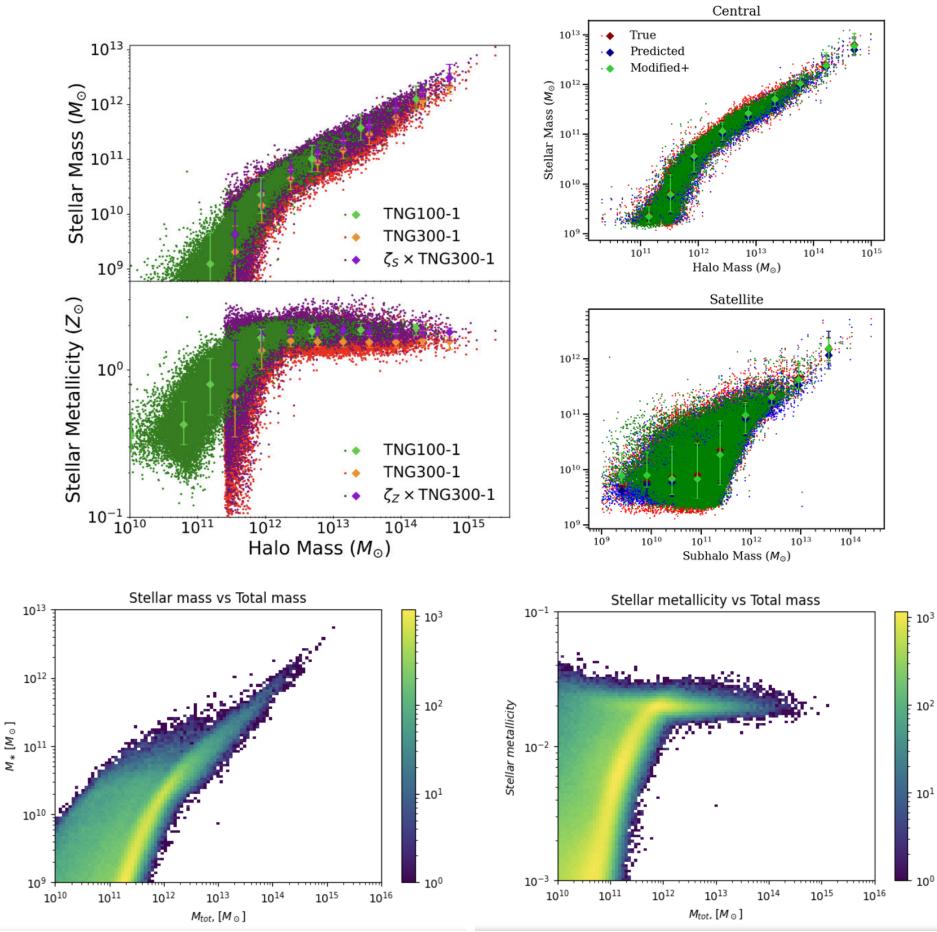


Figure 3: Stellar metallicity and stellar mass vs subhalo mass. Upper left - from the literature [3], upper right - from the literature [4]

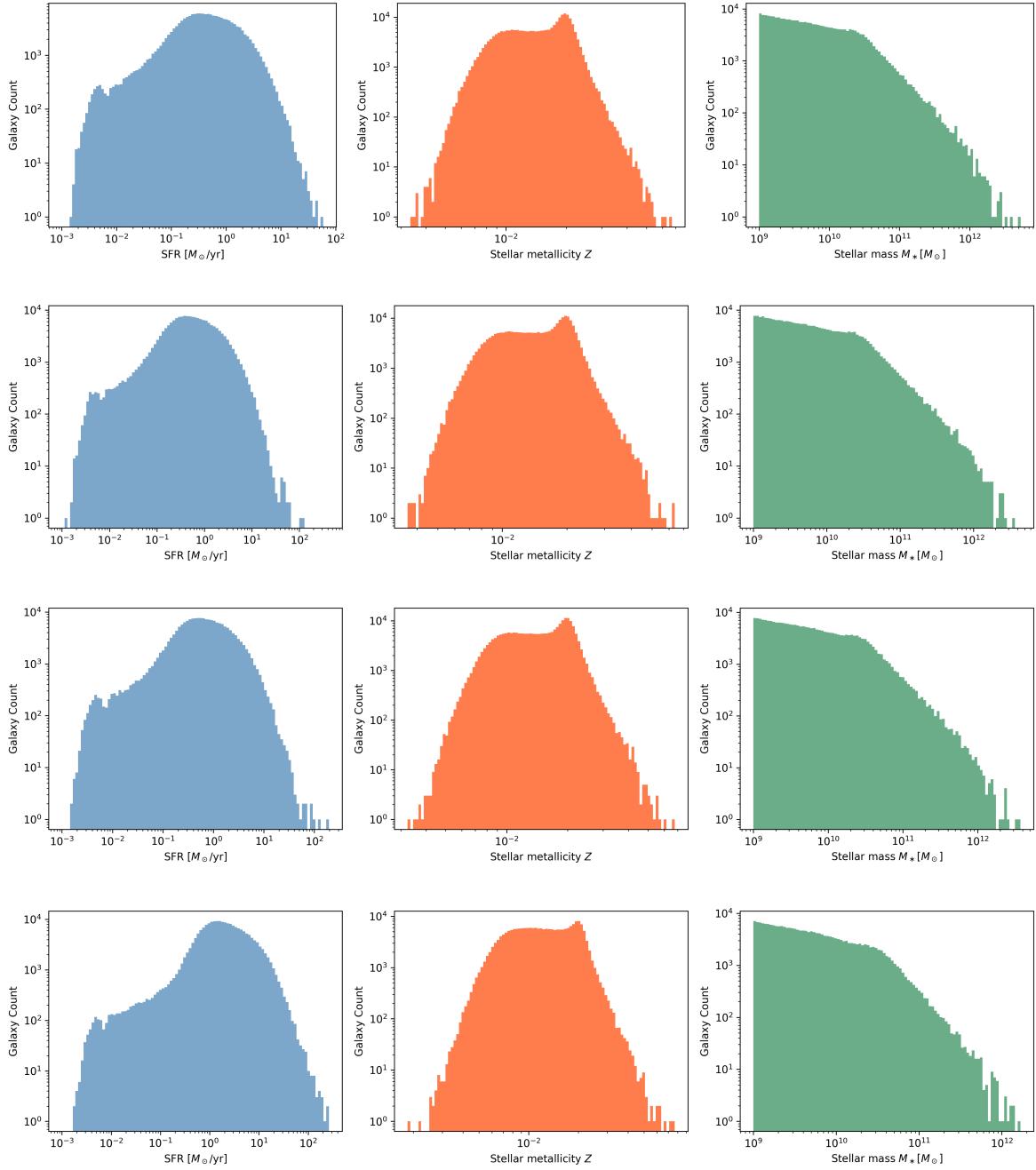


Figure 4: Distributions of values of SFR, metallicity, and stellar mass for redshift  $z = 0.0, 0.1, 0.2, 1.0$  (in order from top to bottom).

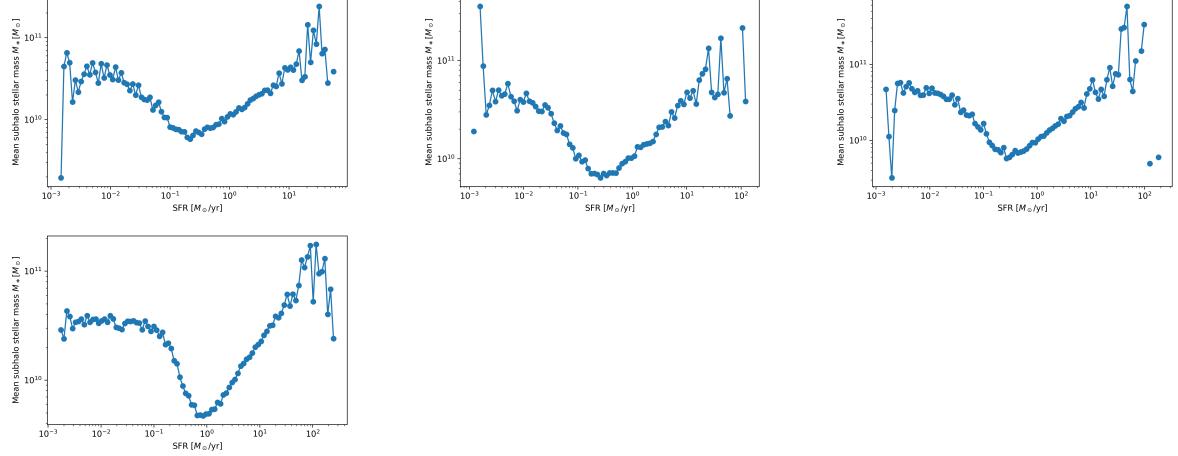


Figure 5: Mean mass dependence on SFR. In the order from left to right, from top to bottom  $z = 0.0, 0.1, 0.2, 1.0$

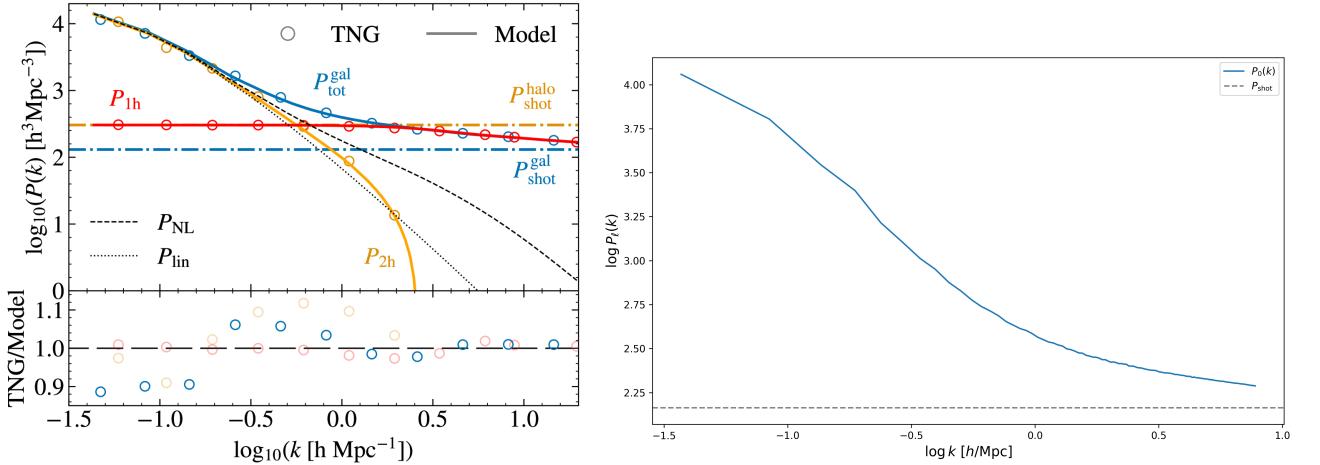


Figure 6: Comparing our SFR-weighted power spectrum (right) to the one from [5] (left) at redshift  $z = 1.0$ . The quantity we want to compare to is the total TNG galaxy power spectrum shown in the upper panel (blue open circles). One difference in our analysis is that we computed shot noise directly from SFR weights, while in [5] shot noise is computed from luminosities.

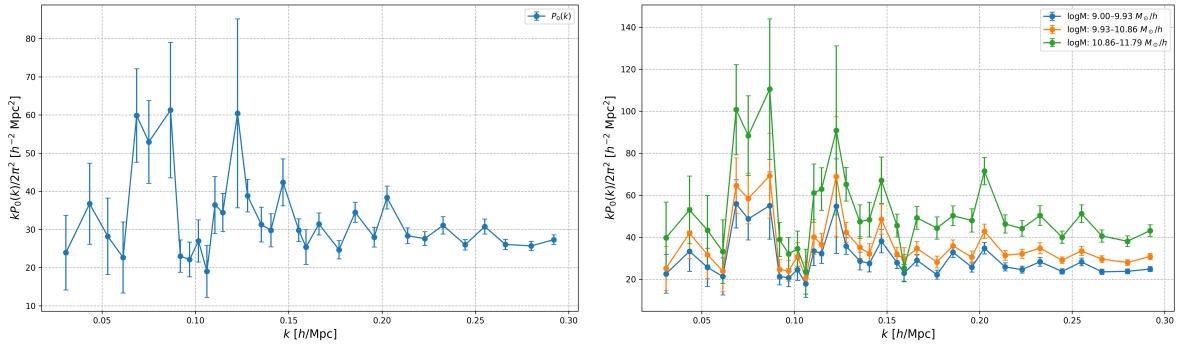


Figure 7: Unweighted power spectrum of subhalos for snapshot  $z = 0$ . Right: overall power spectrum, left: power spectrum for several mass bins. The bin of highest mass is excluded

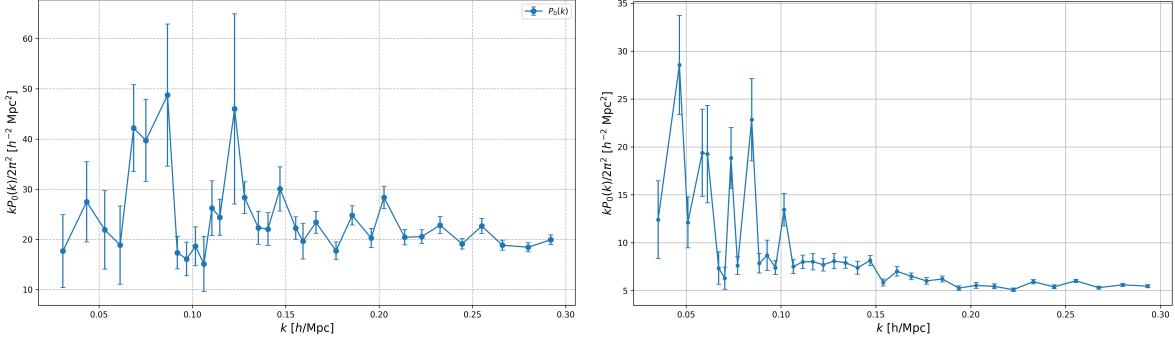


Figure 8: Matter power spectrum for  $z = 0$  obtained through voxelization of dark matter particles (left) and CIC method (right).

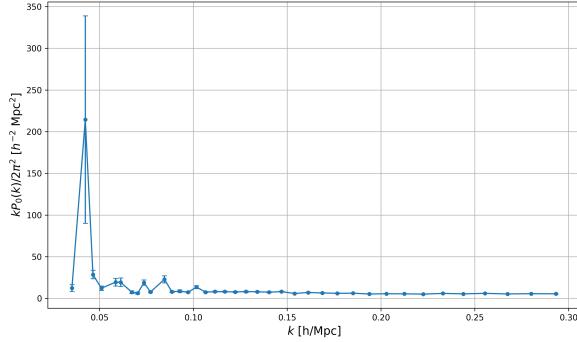


Figure 9: Original matter power spectrum for  $z = 0$  obtained through the CIC method (without large uncertainty bin excluded).

### 3 Matter power spectrum

We used coordinates of Dark matter particles, provided by TNG project, to calculate the matter power spectrum. Due to the large number of particles ( $2500^3$ ), we could not use their coordinates directly to calculate the power spectrum with `pypower`. Instead, we tested two different methods for  $z = 0$ :

- Voxelization. We discretized the simulation box into a uniform 3D grid of size 240 and assigned particles to this cells. Then, we ran `pypower`. We assigned positions to be the mean positions of particles in each voxel and weights to be the total number of particles in each voxel. The resulting power spectrum is presented in the Figure 9.
- Cloud in Cell (CIC). We assigned particle mass to nearby grid points by linearly distributing particle mass among nearest grid points. Then, we calculated the power spectrum using `np.fft.fft`. The value of power spectrum for one of the wavenumber bins had large uncertainty, so we excluded this bin for further analysis. The resulting power spectrum is presented in Figure 9.

With matter and galaxies (subhaloes) power spectra calculated, we calculated galaxy bias. The results are presented in the Figure

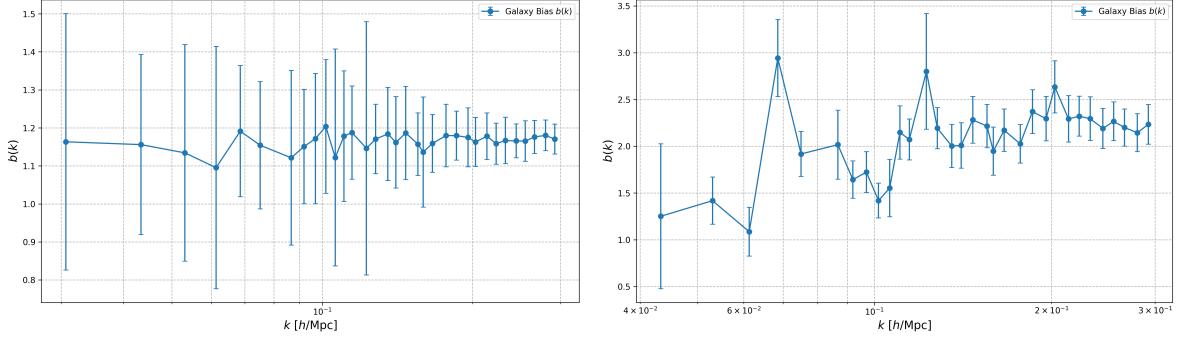


Figure 10: Galaxy bias for  $z = 0$  calculated with matter power spectrum obtained with voxelization (left) and CIC (right)

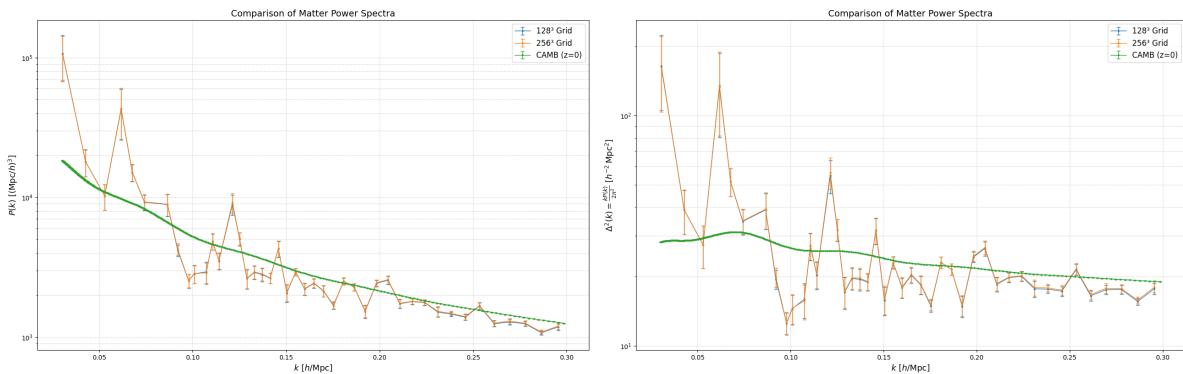


Figure 11: Comparison of Matter Power spectrum of 128 and 256 grid size with CAMB (CIC), scaled by  $h$

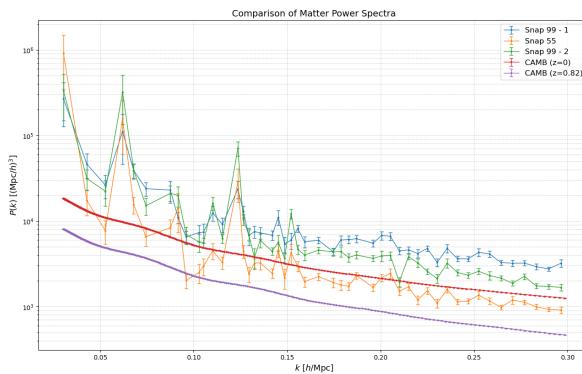


Figure 12: Matter Power spectrum compared with CAMB for different redshifts. The field has been computed using a vectorised algorithm for CIC. Potentially, wraparound issues arise.

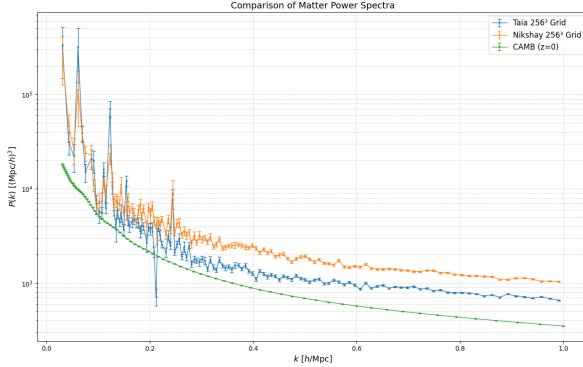


Figure 13: The methods to obtain CIC density field are different.

We then learned about the `Pylians` library [6] for Python which computes CIC density fields far more efficiently than our implementations. Using that to compare with the CAMB nonlinear output, we get:

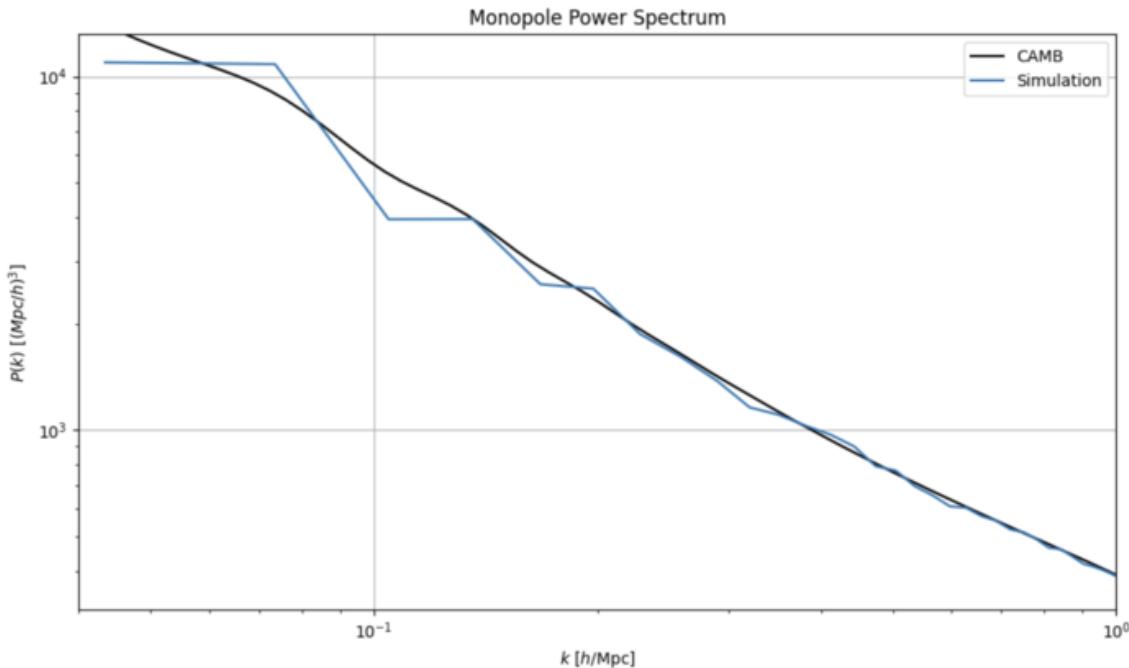


Figure 14: For  $z = 0$ , we obtain this using the dm coordinates from the snapshot and using `Pylians` [6] This is in good agreement, even with the baryon acoustic oscillations.

## 4 Generation of Siren Catalog

We used a simplified version of Dorsa's code to generate Siren Catalogs on Snapshots 50, 55, 60, 65, 70, 75, 80, 85, 90 and 99. These can be mapped to redshift by using the available [GitHub Page](#). This page also contains all the code used in the project. Enjoy!

We used a broken power law

$$\mathbb{P}(\text{host exists in galaxy of mass } M) \propto \begin{cases} M^{1/\delta_l} & M < M_\kappa \\ M^{-1/\delta_h} & M > M_\kappa \end{cases} \quad (1)$$

where  $M_\kappa$  is the pivot mass for the broken power law and  $\delta_l$  and  $\delta_h$  tell us the slopes.

One very interesting effect is thus expected. The nature of dependence of bias on the slopes should change as a function of  $M_\kappa$  depending on whether  $M_\kappa$  is greater than or less than the average mass of the galaxies in the catalog (after the appropriate mass cut). We see this in the results below.

In our study, we applied a mass cut of  $10^9 M_\odot$  to the galaxies and then generated sirens in it. We assigned several properties to the siren catalog: the  $(x, y, z)$  position to the merger, the redshift (through snapshot number), the masses of the merging black holes, the chirp mass, and did this for a set of values in  $\delta_l$ ,  $\delta_h$ , and  $M_\kappa$ .

$$\delta_l = \text{four values linearly taken in } [0.9, 1.5] \quad (2)$$

$$\delta_l = \text{four values linearly taken in } [2, 4] \quad (3)$$

$$M_\kappa = \text{fifteen values logarithmically taken in } [10^9 M_\odot, 10^{12} M_\odot] \quad (4)$$

## 5 Computation of the bias

We used a "linear" fit (this is non-linear, I call it linear because of the  $\mathcal{O}(k)$ ) of the bias given by

$$b(k) = b_0 + b_1 k = \sqrt{P_{GW}/P_{dm}} \quad (5)$$

where  $P_{GW}$  is the power spectrum of the sirens and  $P_{dm}$  is the power spectrum of the dark matter. We minimized the value of  $\chi^2$  over the range of  $[k_{min} = 0.04h/\text{Mpc}, k_{max}]$  to compute the best-fit values of  $b_0$  and  $b_1$ . Plots below show  $b(k = 0.1h/\text{Mpc}) =: b_{GW}$ . In certain cases we also study just  $b_0$  and just  $b_1$ . In plots below, assume  $k_{max} = 1h/\text{Mpc}$  unless specified.

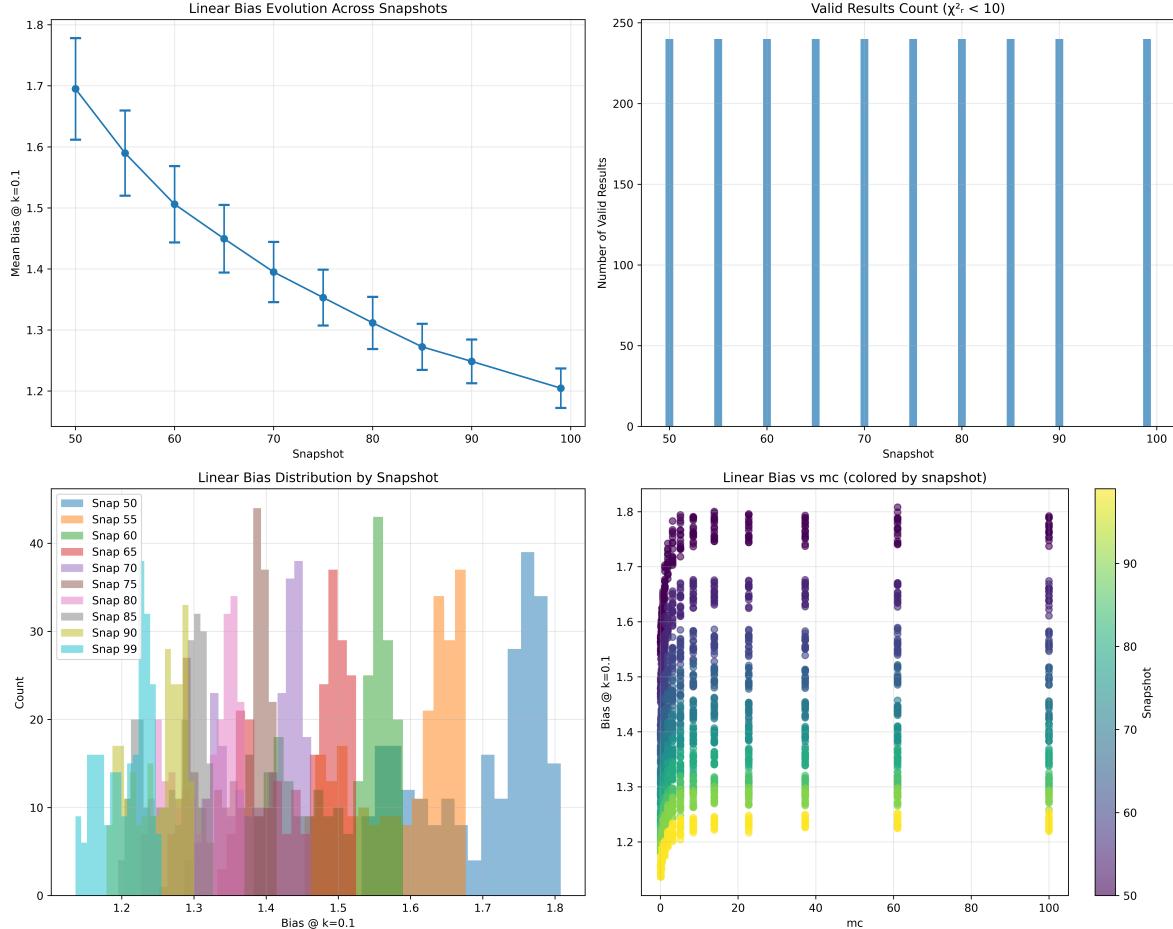


Figure 15: Bias Evolutions

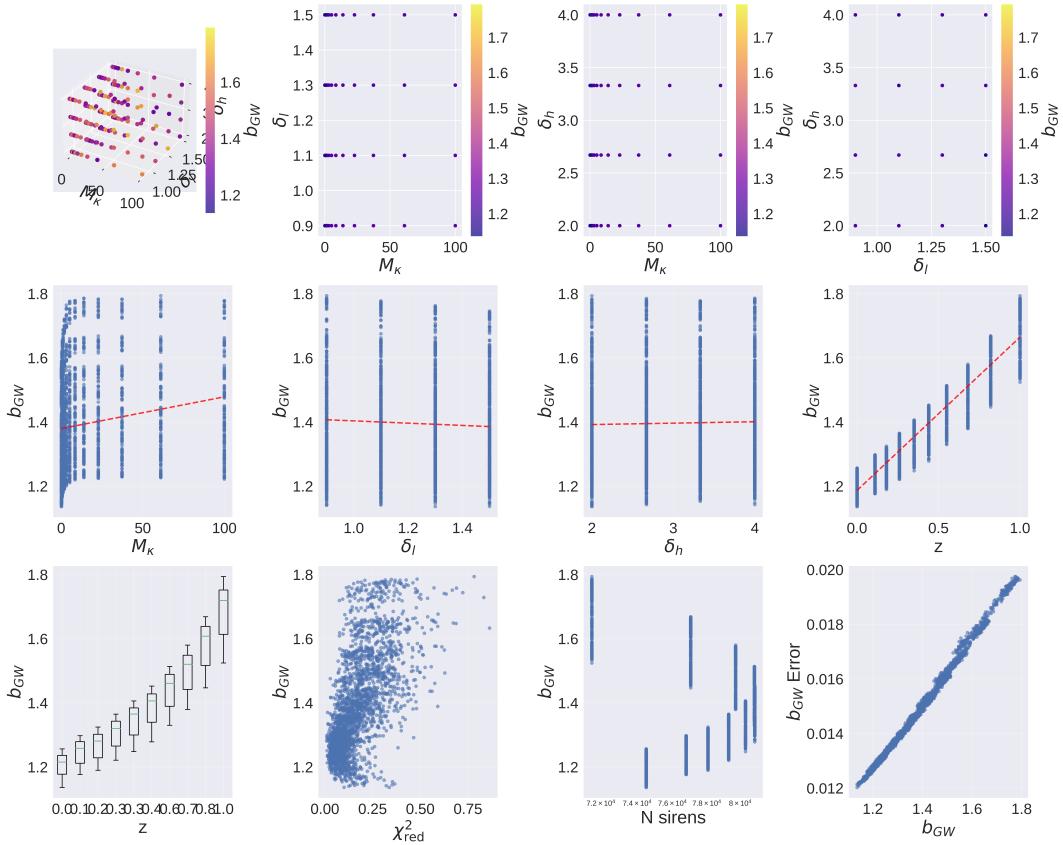
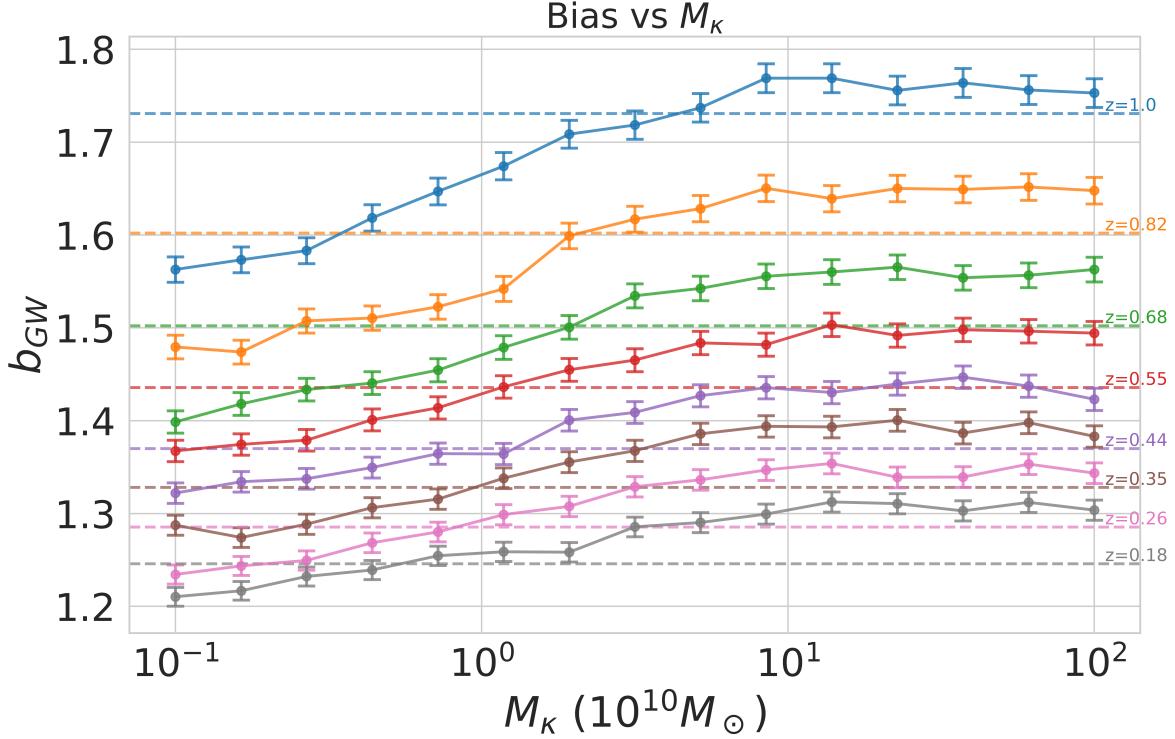


Figure 16: Comprehensive Analysis of the data

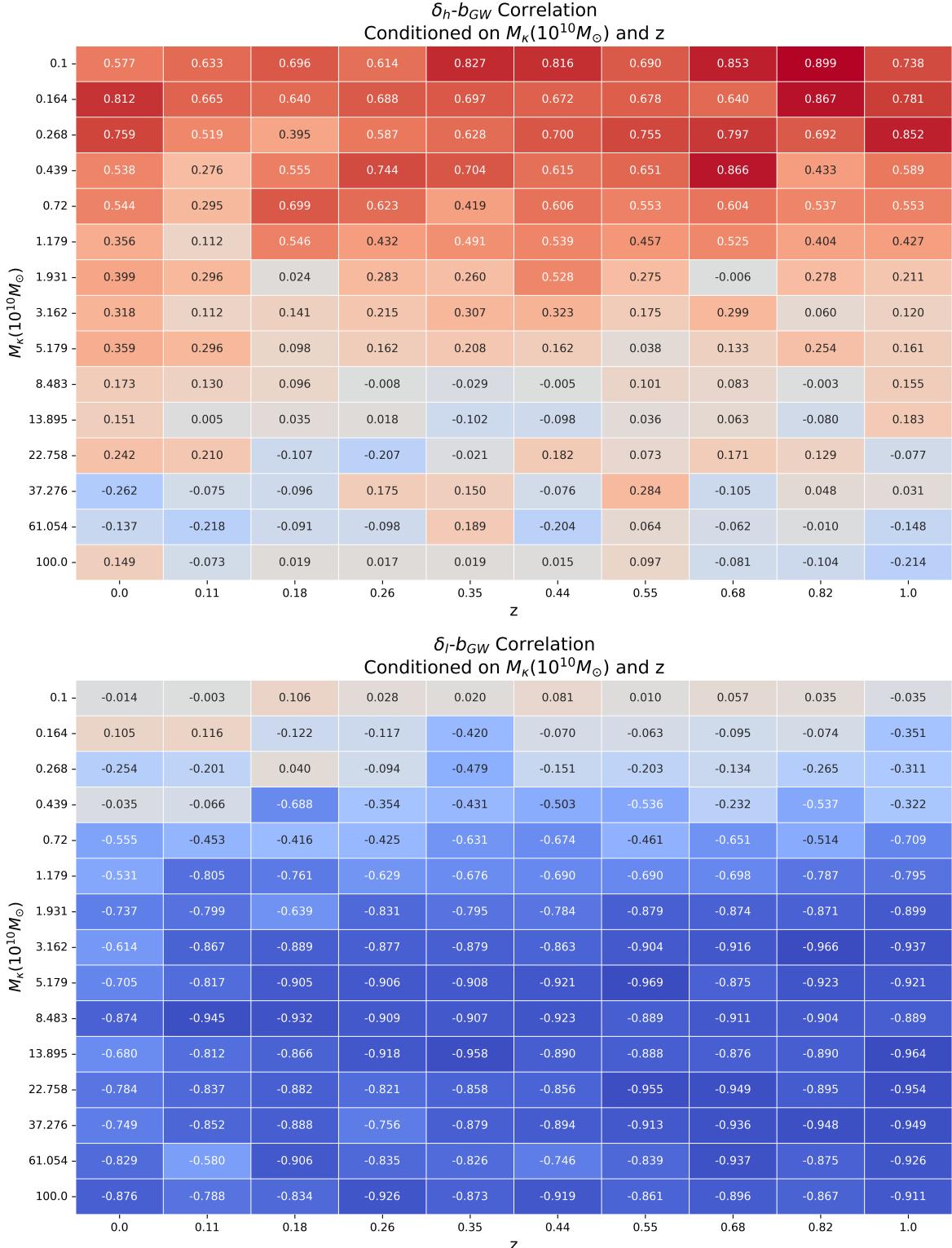


Figure 17: Pearson Correlation Coefficient for  $\delta_{l/h}$  and  $b_{GW}$

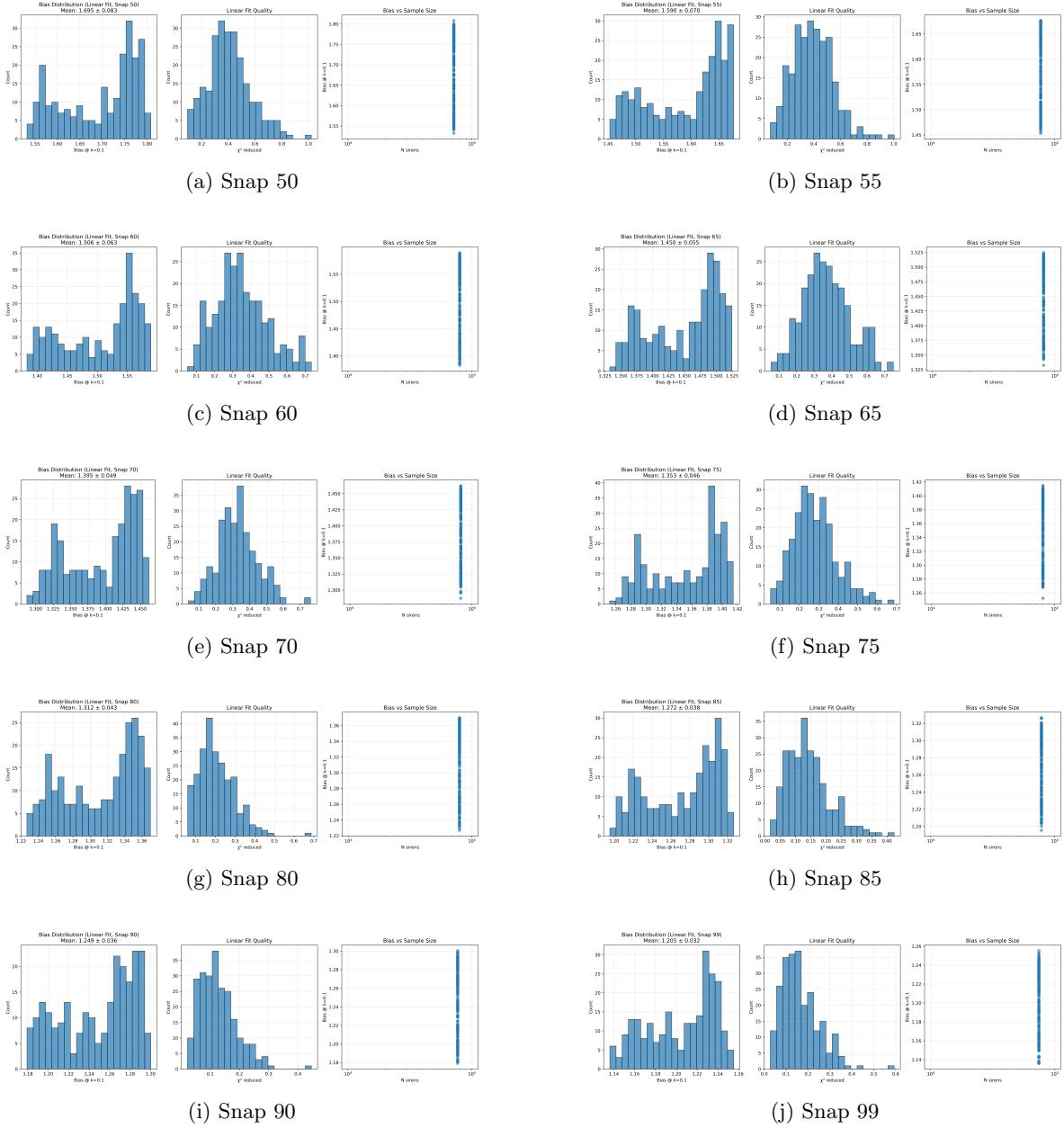


Figure 18: Bias quality metrics for different snapshots across snapshots 50-99.

In the following plots, we reproduce the analysis conducted in Dehghani (2025).

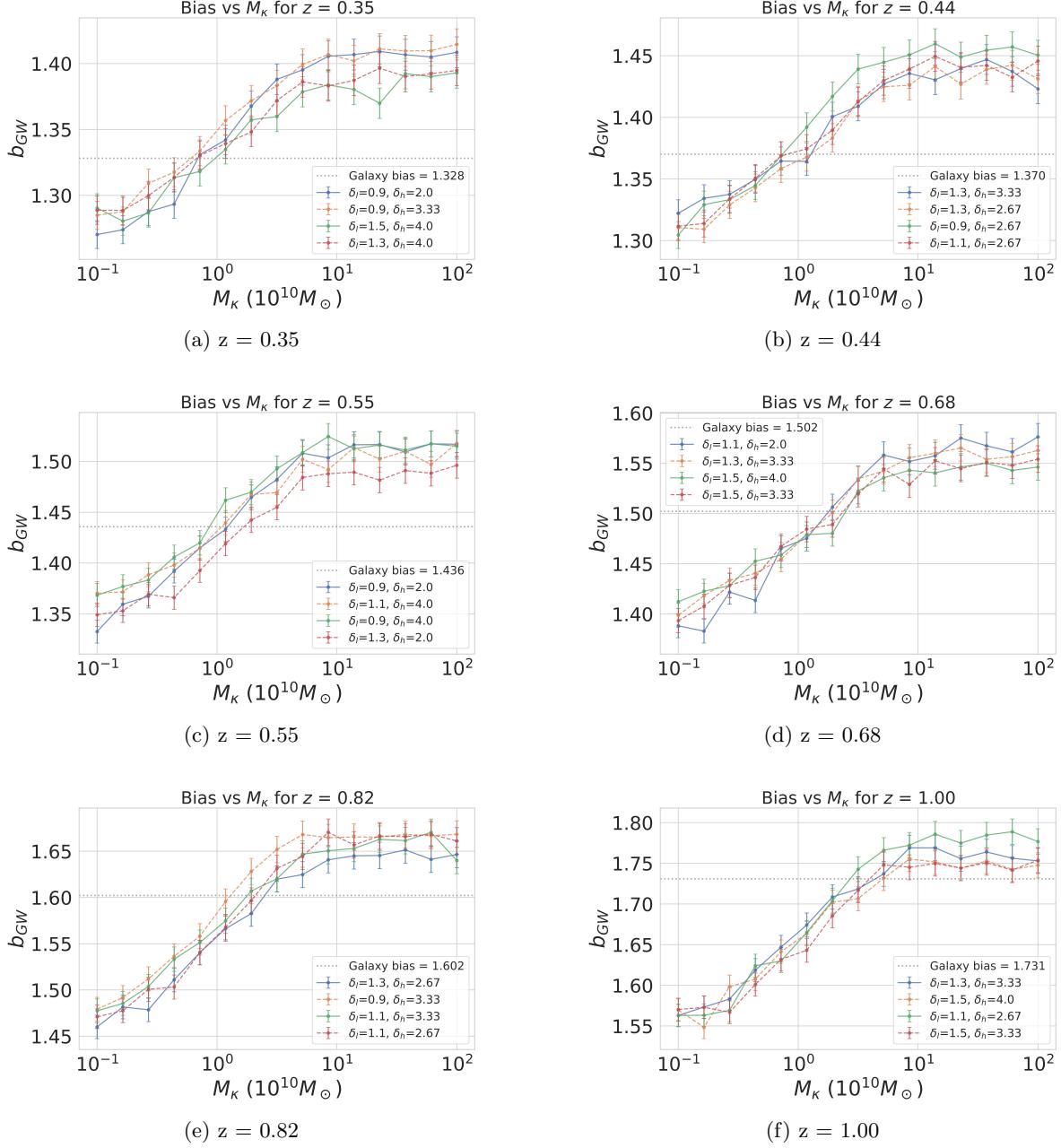


Figure 19:  $b_{GW}$  vs  $M_k$  for different pairs of  $(\delta_l, \delta_h)$  analysis across different redshift values ( $z = 0.35, 0.44, 0.55, 0.68, 0.82, 1.00$ ).

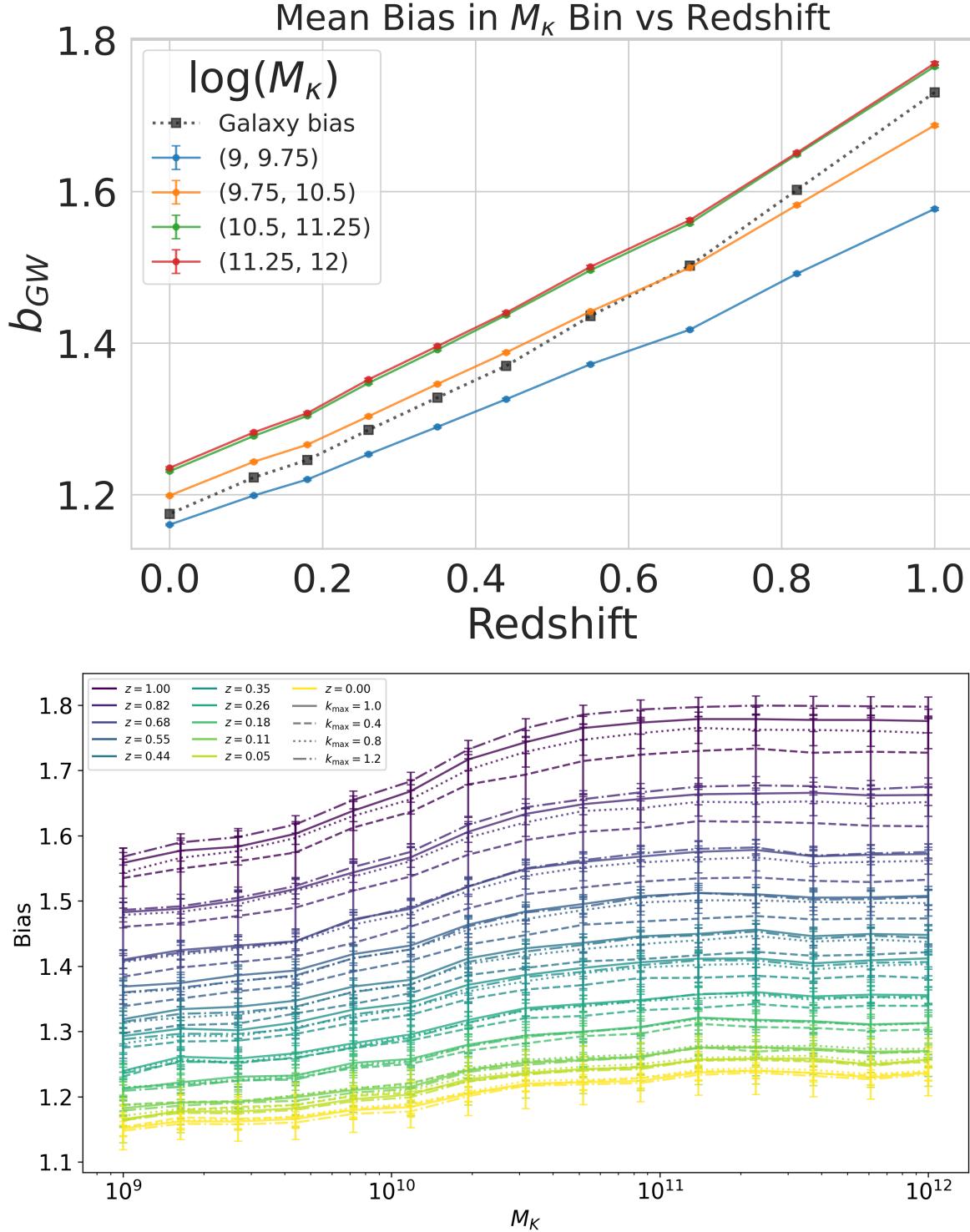


Figure 20:  $b_{GW}$  vs  $M_K$  for different values of  $k_{max}$  while computing best-fit.

## References

- [1] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The *Illustris* simulations: Public data release, 2021.

- [2] Jubee Sohn, Margaret J. Geller, Josh Borrow, and Mark Vogelsberger. Velocity dispersions of quiescent galaxies in *illustiristng*, 2024.
- [3] Harry George Chittenden and Rita Tojeiro. Modelling the galaxy–halo connection with semi-recurrent neural networks. *Monthly Notices of the Royal Astronomical Society*, 518(4):5670–5692, November 2022.
- [4] Jayashree Behera, Rita Tojeiro, and Harry George Chittenden. Optimised neural network predictions of galaxy formation histories using semi-stochastic corrections, 2024.
- [5] Rui Lan Jun, Tom Theuns, Kana Moriwaki, and Sownak Bose. The power spectrum of galaxies from large to small scales: a line-intensity mapping perspective, 2025.
- [6] Francisco Villaescusa-Navarro. Pylians: Python libraries for the analysis of numerical simulations. Astrophysics Source Code Library, record ascl:1811.008, November 2018.