

Quantifying the impact of AGN feedback on the large-scale matter distribution using two- and three-point statistics

Bipradeep Saha,¹* Sownak Bose,² †

¹*Department of Physical Sciences, IISER - Kolkata, India*

²*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Feedback from active galactic nuclei (AGN) plays a critical role re-distributing matter on scales larger than individual galaxies. With several present and upcoming surveys like DES, LSST, and *Euclid*, tasked with quantifying the matter distribution accurately, a comprehensive understanding of these effects is vital. Hydrodynamical simulations are essential tools for this purpose. However, models of AGN feedback are subject to a range of approaches and parameterisations, making it important to consider the effect of varying different parameters. We use the EAGLE simulations to investigate how changing the subgrid viscosity and heating temperature of AGN affects the matter distribution in simulations. In particular, we use 2-point correlation functions (2pCF) and cumulants of the matter distribution to quantify the effect of changing these parameters, and find that changing the viscosity has very negligible change, $\approx 10\%$ on scales larger than galaxies i.e. $> 1h^{-1}$ Mpc. Significant changes appear upon varying the heating temperature of the AGN, with differences up to 70% in the clustering of gas on small scales ($\lesssim 1h^{-1}$ Mpc). By considering the suppression of the power spectrum, alongside the black hole accretion rate density in these simulations, we identify the range $z = 1.5 - 1$ as the epoch when AGN feedback starts to dominate in the EAGLE model. Finally, we also quantify the effect of varying the AGN subgrid parameters on the gas distribution using the 3-point correlation function (3pCF), and find that it provides complementary information to the 2pCF, where different models vary more substantially on the scale of individual haloes, but have comparable effects on larger scales.

Key words: large-scale structure of Universe – hydrodynamics – methods: statistical – methods: numerical

1 INTRODUCTION

Cosmological simulations have greatly improved our understanding of the physics of galaxy formation and are widely used to guide the interpretation of observations and the design of new observational campaigns and instruments. Simulations are useful for testing how different physical processes affect galaxy formation, which can then be compared with observations in order to develop a more complete understanding of the Universe.

In the following decades, ambitious observational campaigns will aim to pin down the source of the accelerated expansion of the Universe – whether it is a cosmological constant, an exotic source of dark energy, or resulting from modifications to General Relativity – as well as the impact of the neutrinos in the Universe, and detailed statistics of primordial fluctuations to test for deviations from Gaussianity. To tackle these challenges, the scientific community will perform far more precise surveys of the Universe on large scales. The next generation of galaxy surveys will rely on different proxies of matter in the Universe to answer the questions raised above. Previous and ongoing surveys like the Sloan Digital Sky Survey (SDSS, Huff et al. 2014), the Kilo Degree Survey (KiDS, de Jong et al. 2013), and the

Dark Energy Survey (DES, Troxel et al. 2018) have demonstrated the enormous potential of this era of ‘precision cosmology’. Many experiments are planned to obtain better constraining power on the cosmological model, such as the Vera C. Rubin Observatory’s Large Synoptic Survey Telescope (LSST, Ivezić et al. 2019) and *Euclid* (Laureijs et al. 2011). Before inferring cosmological parameters from these surveys, we need predictions of the theoretical matter power spectrum, $P(k)$, which quantifies the amount of statistical power in a given Fourier mode of the matter over-density field. Previous work suggests that the $P(k)$ needs to be modeled to a precision such that ΔP is constrained to being $\approx 1\%$ at $k_{\max} = 10h/\text{Mpc}$ or even larger scales (e.g. Huterer & Takada 2005; Hearin et al. 2012).

While it is possible to model the total matter power spectra via (semi-)analytical methods (Takahashi et al. 2012) or by using dark matter-only simulations, it is now increasingly clear that the baryonic process could significantly impact the distribution of matter, and that these effects need to be considered while analyzing the data from the next-generation surveys (van Daalen et al. 2011; Hellwing et al. 2016; Chisari et al. 2018). Thus a crucial improvement is required in modeling the large-scale structure, i.e., the description of the impact of galaxy formation on the distribution of matter. Processes that heat and cool the gas, re-distribute it or transform it into stars have to be included in the models. The main effect to include is the suppression of power at the scales of a few Mpc, which is associated with the

* E-mail: bipradeepsaha04@gmail.com

† E-mail: sownak.bose@durham.ac.uk

gas ejected by Active Galactic Nuclei (AGN). A range of “effective” models to account for this problem have been suggested (e.g. Mead et al. 2015; Schneider & Teyssier 2015). However, state-of-the-art hydrodynamical simulations, which model the co-evolution of the dark and ordinary matter self-consistently, are the most effective means to tackle this problem (e.g. Vogelsberger et al. 2014; Springel et al. 2017). The success of these techniques depends on the flexibility of these models to capture the true underlying matter distribution. The most significant limiting factor for hydrodynamical simulations is that these they are computationally very expensive compared to the dark matter-only simulations and are more difficult to control, as the multitude of parameters used to describe baryonic processes and feedback are poorly constrained.

In this work, we investigate how varying the sub-grid physics for AGN affects the distribution of gas and thus the total matter power spectra. We use the Virgo Consortium’s EAGLE Project (Evolution and Assembly of GaLaxies and their Environments, Schaye et al. 2015), which is a suite of cosmological smooth particle hydrodynamic simulations of the Λ Cold Dark Matter (Λ CDM) Universe. The main models were run in volumes of 25 – 100 co-moving Mpc (cMpc) and the resolution is sufficient enough to marginally resolve the Jeans scales in the warm ($T \sim 10^4$ K) interstellar medium (ISM). The aim of the present study is to help us understand how the physics of the subgrid manifest the large scale matter distribution and identify the statistics that are most responsive to these changes, at least qualitatively.

In order to assess the effect of AGN feedback on the matter distribution, we make use of the two-point correlation function (2pCF) and the power spectra of different matter components in a given simulation and compare them for different variations in the sub-grid model. We then compute the cross-correlation between the gas field and the black-hole field weighted by the instantaneous accretion rate to quantify the specific impact of the growth and evolution of black holes specifically. This enables us to constrain the both the spatial and temporal scales where AGN feedback start to impact the large-scale matter distribution. To quantify this distribution further, we consider the cumulants of the matter over-density field. Finally, to quantify the non-Gaussian nature of the overdensity field and the baryonic impacts on it, we investigate the three-point correlation function (3pCF) of the gaseous component of the matter distribution.

This manuscript is organized as follows. Section 2 describes the simulations and the model variations that we use. In Section 3, we describe the methods that we use for our analysis. In Section 4, we present our main results and explain the different features we observe, followed by a discussion and a summary of our main conclusions in Sections 5 and 6, respectively.

2 SIMULATION DETAILS

For our primary dataset, we use the Virgo Consortium’s EAGLE project. EAGLE makes use of the smooth particle hydrodynamics (SPH) method run with a modified version of the Tree-PM SPH code GADGET3, last described by Springel (2005). The main modifications are the formulation of SPH, the time stepping, and, most importantly, the subgrid physics.

The simulation was calibrated to match the relation between stellar mass and halo mass, galaxies’ present-day stellar mass function, and galaxy sizes (Schaye et al. 2015; Crain et al. 2015). The subgrid physics used in EAGLE is based on that developed for OWLS (Schaye et al. 2010) and COSMO-OWLS (Le Brun et al. 2014). It includes element-by-element radiative cooling for 11 elements, star formation,

stellar mass-loss, energy feedback from star formation, gas accretion onto and mergers of supermassive BHs (black holes), and AGN feedback. However, there are several changes from OWLS; the most important ones are implementations of energy feedback from star formation (which is thermal rather than kinetic), the accretion of gas onto BHs (which accounts for angular momentum), and the star formation law (which depends on metallicity). As they are of primary importance to our present work, we describe the AGN feedback model employed in EAGLE in the following subsections.

2.1 Gas accretion onto black holes

The rate at which BHs accrete gas depends on the mass of the BH, the local density and temperature of the gas, the velocity of the BH relative to the ambient gas, and the angular momentum of the gas with respect to the BH. Specifically, the gas accretion rate, \dot{m}_{accr} , is given by the minimum of the the Eddington rate:

$$\dot{m}_{\text{Edd}} = \frac{4\pi G m_{\text{BH}} m_p}{\epsilon_r \sigma_T c} \quad (2.1)$$

and

$$\dot{m}_{\text{accr}} = \dot{m}_{\text{Bondi}} \times \min \left(C_{\text{Visc}}^{-1} (c_s/V_\phi)^3, 1 \right), \quad (2.2)$$

where \dot{m}_{Bondi} is the Bondi & Hoyle rate for spherically symmetric accretion (Bondi & Hoyle 1944). It is given by :

$$\dot{m}_{\text{Bondi}} = \frac{4\pi G^2 m_{\text{BH}}^2 \rho}{(c_s^2 + v^2)^{3/2}}. \quad (2.3)$$

The mass growth rate of the BH is given by

$$\dot{m}_{\text{BH}} = (1 - \epsilon_r) \dot{m}_{\text{accr}}. \quad (2.4)$$

In the above equations, m_p is the proton mass, σ_T is the Thomson cross-section, c is the speed of light, $\epsilon_r = 0.1$ is the radiative efficiency, v is the relative velocity of the BH and the gas and finally, V_ϕ is the rotation speed of the gas around the BH computed using Equation 16 of Rosas-Guevara et al. (2015). Here, C_{Visc} is a free parameter related to the viscosity of a notional subgrid accretion disc. The factor $(c_s/V_\phi)^3 / C_{\text{Visc}}$ by which the Bondi rate is multiplied in Equation 2.2 is equivalent to the ratio of the Bondi and viscous time scales (see Rosas-Guevara et al. 2015). The critical ratio of V_ϕ/c_s above which angular momentum is assumed to suppress the accretion rate scales as $C_{\text{Visc}}^{-1/3}$. Thus, larger values of C_{Visc} correspond to a lower subgrid kinetic viscosity, and so act to delay the growth of BHs by gas accretion and, by extension, the onset of quenching by AGN feedback.

2.2 AGN feedback

AGN feedback in the EAGLE simulation is implemented thermally and stochastically. By making the feedback stochastic, one can control the amount of energy per feedback event even if the mean energy injected per unit mass is fixed. The energy injection rate is given by $\epsilon_f \epsilon_e \dot{m}_{\text{accr}} c^2$, where ϵ_f is the fraction of the radiated energy that couples with the interstellar medium (ISM). The value of ϵ_f needs to be chosen by calibrating to the observation. In Schaye et al. (2015) it is justified that $\epsilon_f = 0.15$ and $\epsilon_r = 0.1$ is a suitable choice of these parameters.

Each BH carries a reservoir of feedback energy, E_{BH} . After each time step Δt , energy equivalent to $\epsilon_f \epsilon_r \dot{m}_{\text{accr}} c^2 \Delta t$ is added to the reservoir. Once a BH has stored sufficient energy to heat at least one

Model Name	Box size [h ⁻¹ Mpc]	N	C _{visc} /2π	ΔT _{AGN} log ₁₀ [k]
Ref	33.885	752 ³	10 ⁰	8.5
ViscHi	33.885	752 ³	10⁻²	8.5
ViscLo	33.885	752 ³	10²	8.5
AGNdT8	33.885	752 ³	10 ⁰	8
AGNdT9	33.885	752 ³	10 ⁰	9
NoAGN	33.885	752 ³	-	-

Table 1. Table summarizing the parameter variation between different runs of the EAGLE simulation. Here N is the total number of particles of each type (DM and Gas), C_{visc} is the viscosity parameter and ΔT_{AGN} is the AGN heating parameter. The bold quantities indicate the parameters that were changed with respect to the reference values for that model variation.

fluid element of mass m_g , the BH is allowed to heat each of its SPH neighbours by a temperature ΔT_{AGN} , stochastically. For each neighbour, the heating probability is given by:

$$P = \frac{E_{\text{BH}}}{\Delta \epsilon_{\text{AGN}} N_{\text{ngb}} \langle m_g \rangle} \quad (2.5)$$

where, ϵ_{AGN} is the change in internal energy per unit mass corresponding to the temperature increment, N_{ngb} is the number of gas neighbours of the BH and $\langle m_g \rangle$ is their mean mass.

Larger values of ΔT_{AGN} yield more energetic feedback events, generally resulting in reduced radiative losses (Crain et al. 2015). However, larger values also make the feedback more intermittent. In general, the ambient density of gas local to the central BH of galaxies is greater than that of star-forming gas distributed throughout their discs, so a higher heating temperature is required to minimise numerical losses.

In this work, we used two different model variations from the EAGLE run:

- ViscHi and ViscLo
- AGNdT8 and AGNdT9

which correspond to variations in the subgrid physics parameters relevant to the processes described above. Their details are summarized in Table 1. As mentioned above, these model variations affect gas accretion onto the black holes and the effective strength of AGN feedback, respectively (Crain et al. 2015; Schaye et al. 2015).

3 METHODS

This section introduces the definition of various statistical tools used to characterise the distribution of the matter density field in our analysis.

If $\rho(\mathbf{x})$ is the matter density at the point \mathbf{x} , and $\bar{\rho}$ is the mean matter density of the universe, we define the density contrast as:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1.$$

Then, the Fourier modes for the density contrast field for the set of N particles with mass m_i in a periodic box of length L of volume V_u is defined as (Peebles 1980; Mo et al. 2010):

$$\delta_{\mathbf{k}} = \frac{1}{M} \sum_{i=1}^N m_i \exp(-i\mathbf{x} \cdot \mathbf{k}) = \frac{1}{V_u} \int \delta(\mathbf{x}) \exp(-i\mathbf{x} \cdot \mathbf{k}) \quad (3.1)$$

where $\sum m_i = M$ and the second equality is in the continuum limit.

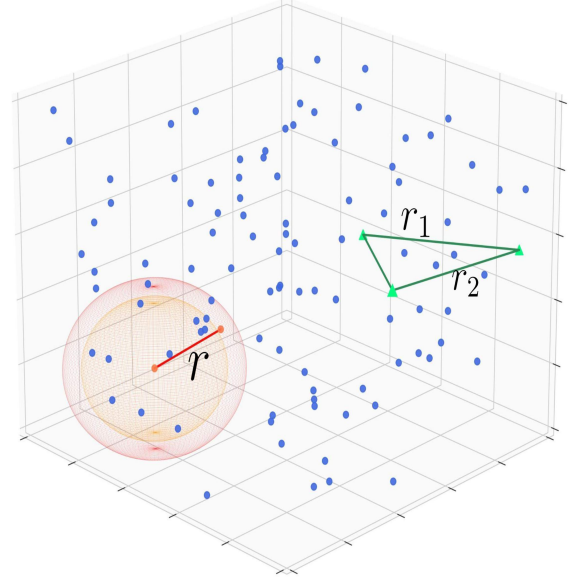


Figure 1. Illustration on how to measure 2pCF and 3pCF. For the 2pCF we sit at a point and compute the number of point inside a spherical shell of radius r , and thickness dr . We then spatially average this across the data to get the DD (data-data) pair counts. We then repeat the same procedure for random data set and get RR (random-random) pair counts. Similarly for 3pCF we count the triangles having side r_1 and r_2 in the observed and the random data set to get DDD and RRR value. Then using the generic estimators we can estimate the 2pCF and 3pCF.

3.1 Spatial Correlation Functions

N -point statistics are essential tools for quantifying a distribution of points in a field. In cosmology, correlation functions are used to quantify the clustering of objects in the Universe, test hierarchical scenarios for structure formation, test Gaussianity of the initial conditions, and test various models for the clustering bias between luminous and dark matter.

The *two-point correlation function* (2pCF) measures the excess probability of finding two correlated points separated by distance r (Peebles 1980):

$$\delta P = n^2 [1 + \xi(r)] dV_1 dV_2, \quad (3.2)$$

where δP is the joint probability of finding particles in volume element dV_1 and dV_2 separated by distance r , n is the mean number density of tracers and $\xi(r)$ is the 2pCF. Its Fourier transform, known as the *power spectrum*, $P(k)$, is given by:

$$P(k) = V_u \langle |\delta_{\mathbf{k}}|^2 \rangle = V_u \langle \delta_{\mathbf{k}} \delta_{-\mathbf{k}} \rangle \quad (3.3)$$

$$\xi(r) = \int P(k) e^{-i\mathbf{k} \cdot \mathbf{x}} d^3\mathbf{x} = \frac{1}{2\pi^2} \int P(k) \frac{\sin(kx)}{kx} d^3\mathbf{x} \quad (3.4)$$

where $\langle \cdot \rangle$ denotes the ensemble average, and $\delta_{\mathbf{k}}$ is as defined in Equation 3.1.

Similarly the *three-point correlation function* (3pCF) measures the excess probability of finding three correlated points (i.e., triangular configurations, Peebles 1980):

$$\delta P = n^3 [1 + \xi(r_a) + \xi(r_b) + \xi(r_c) + \zeta(r_a, r_b, r_c)] dV_1 dV_2 dV_3, \quad (3.5)$$

where the terms inside $[\cdot]$ is the full 3pCF and $\zeta(r_a, r_b, r_c)$ is the reduced 3pCF and r_a, r_b, r_c are the three sides of the triangle formed by the three points.

Figure 1 shows visual depictions of the 2- and 3-point correlation functions. Measuring the spatial correlation function from the above definitions is computationally expensive. Therefore we use estimators for computing each correlation function:

- For computing the 2pCF and $P(k)$ we use the FFT based approach following the definitions in Equations 3.3 and 3.4. In practice, we use the code by [Villaescusa-Navarro \(2018\)](#) for this purpose.
- For computing the 3pCF we use a Legendre polynomial decomposition described by [Philcox et al. \(2021\)](#), which uses the generic estimator of the 3pCF:

$$\zeta(r_a, r_b, r_c) = \frac{NNN}{RRR} \quad (3.6)$$

as the primary definition, but uses Legendre polynomials to speed up the counting of number of triangles. Here, N is defined as, $N := D - R$. Hence, Equation 3.6 contains terms like DDD , RRR , DRR , DDR , where DDD is number of triangles from the data set with sides $r_1 \pm \Delta r$ and $r_2 \pm \Delta r$, RRR is the number of triangles from the random data set with same side lengths. For DRR we mix one part of the data with two parts of random data, and then compute the number of triangles with sides $r_1 \pm \Delta r$ and $r_2 \pm \Delta r$, and vice versa for DDR counts. The algorithm sits on each point in the dataset, and computes the spherical harmonic expansion of the density field in concentric spherical shells (radial bins) around that point, which is then combined to yield the multipole moments around this point, and then translation average is taken to yield, ζ_ℓ . Using this method, the 3pCF is given by:

$$\zeta(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\ell} \frac{\sqrt{2\ell+1}}{4\pi} (-1)^\ell \zeta_\ell(r_1, r_2) L_\ell(\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2) \quad (3.7)$$

where, $\mathbf{r}_1, \mathbf{r}_2$ parameterize the triangle, ζ_ℓ are re-scaled Legendre Polynomials ([Slepian & Eisenstein 2015](#)) and $L_\ell(x)$ are Legendre polynomials of order ℓ . For our purpose, we set $\ell_{max} = 5$, depending on the computational resources available and triangle configurations we are interested in.

3.2 Cumulants of Matter Distribution

The non-linear evolution of the density field, δ , drives the field and its distribution away from an initially Gaussian distribution ([Bernardeau 1992](#)). One way to study these deviations is by using cumulants or reduced moments ([Fry 1984, 1985](#)). For a Gaussian field, the first two central moments are sufficient to characterize the full distribution. Thus, higher order cumulants of the density field are useful to characterize the non-linear density field in the presence of galaxy formation.

The n -th cumulant of the density field, δ , is defined by a recursive relation to the n -moment. It is expressed by the cumulant generating function, $K(\delta)$, as:

$$\langle \delta^n \rangle := \frac{\partial^n K(\delta)}{\partial t^n} = \frac{\partial^n \ln \langle e^{t\delta} \rangle}{\partial t^n}. \quad (3.8)$$

In our analysis we consider cumulants up to 5th order. In terms of

central moments, they are given by the following equations:

$$\begin{aligned} \langle \delta^1 \rangle_c &= 0 \text{ (mean)} \\ \langle \delta^2 \rangle_c &= \langle \delta^2 \rangle \equiv \sigma^2 \text{ (variance)} \\ \langle \delta^3 \rangle_c &= \langle \delta^3 \rangle \text{ (skewness)} \\ \langle \delta^4 \rangle_c &= \langle \delta^4 \rangle - 3 \langle \delta^2 \rangle_c^2 \text{ (kurtosis)} \\ \langle \delta^5 \rangle_c &= \langle \delta^5 \rangle - 10 \langle \delta^3 \rangle_c \langle \delta^2 \rangle_c, \end{aligned}$$

where the subscript c denotes cumulants.

Our analysis is based on the fact that the primordial matter density field is almost Gaussian. To remain unaffected by the impact of local maxima and minima in the density field, we first smooth the density field with a spherical filter of some size; this indeed degrades the resolution of the simulated density field to one that may be more readily calculated in observations. In our analysis, we also study how the cumulants change with smoothing scale chosen.

4 RESULTS

4.1 Comparing different models

Before we look into the quantitative differences between the different models, we first look at the qualitative differences. In Figure 2, we look at the temperature fields of the gaseous component for the different models at $z = 0$ projected along the z -axis. The first row shows the gas temperature fields for AGNdT8, 9 model, while the bottom row shows the difference in the temperature fields for AGNdT8-9 and ViscHi-Lo in a 10 Mpc h^{-1} region centred on the most massive halo in the box. Note that we only show the temperature difference field for the comparison of ViscHi/Lo model to highlight their subtle differences more clearly.

From the first row, we notice that gas in AGNdT9 is hotter as well as more dispersed compared to the AGNdT8 model i.e when one looks at the bubbles in the same region, for e.g upper right corners, bubbles in AGNdT9 occupy more area compared to AGNdT8. Additionally, the black, empty regions are less prominent in AGNdT9 compared to AGNdT8 which implies hotter gas is dispersed over larger regions in AGNdT9 compared to AGNdT8. Even in the zoomed region around the most massive halo, we see that gas in AGNdT9 is typically hotter. We observe that near the centre of the halo, gas in the AGNdT8 is hotter than AGNdT9 while, as we move away from the center, AGNdT9 is hotter than AGNdT8 as AGNdT9 has higher heating temperature than AGNdT8, which leads to higher energy injection into the surrounding circumgalactic medium (CGM). Due to this higher energy injection, we expect gas heated by AGN to be transported further in AGNdT9 model. Hence, we see cooler gas in AGNdT9 compared to AGNdT8 around the centre of the halo, but as we move further, the temperature of the gas increases more rapidly in AGNdT9. This also provides initial indication that gas in AGNdT9 is less clustered than AGNdT8.

Next, we look at the zoomed region for ViscHi-Lo. In ViscLo, subgrid viscosity is lowered, which results in the delayed onset of AGN feedback. As result more time is available for the gas to cool down in the CGM. Moreover, the energy injection by BHs in the ViscLo model is lowered compared to the ViscHi model. Therefore, we expect the matter in ViscLo model to be more clustered, at least on the small scales compared to ViscHi.

Now we look at a quantitative analysis of the gas distributions in

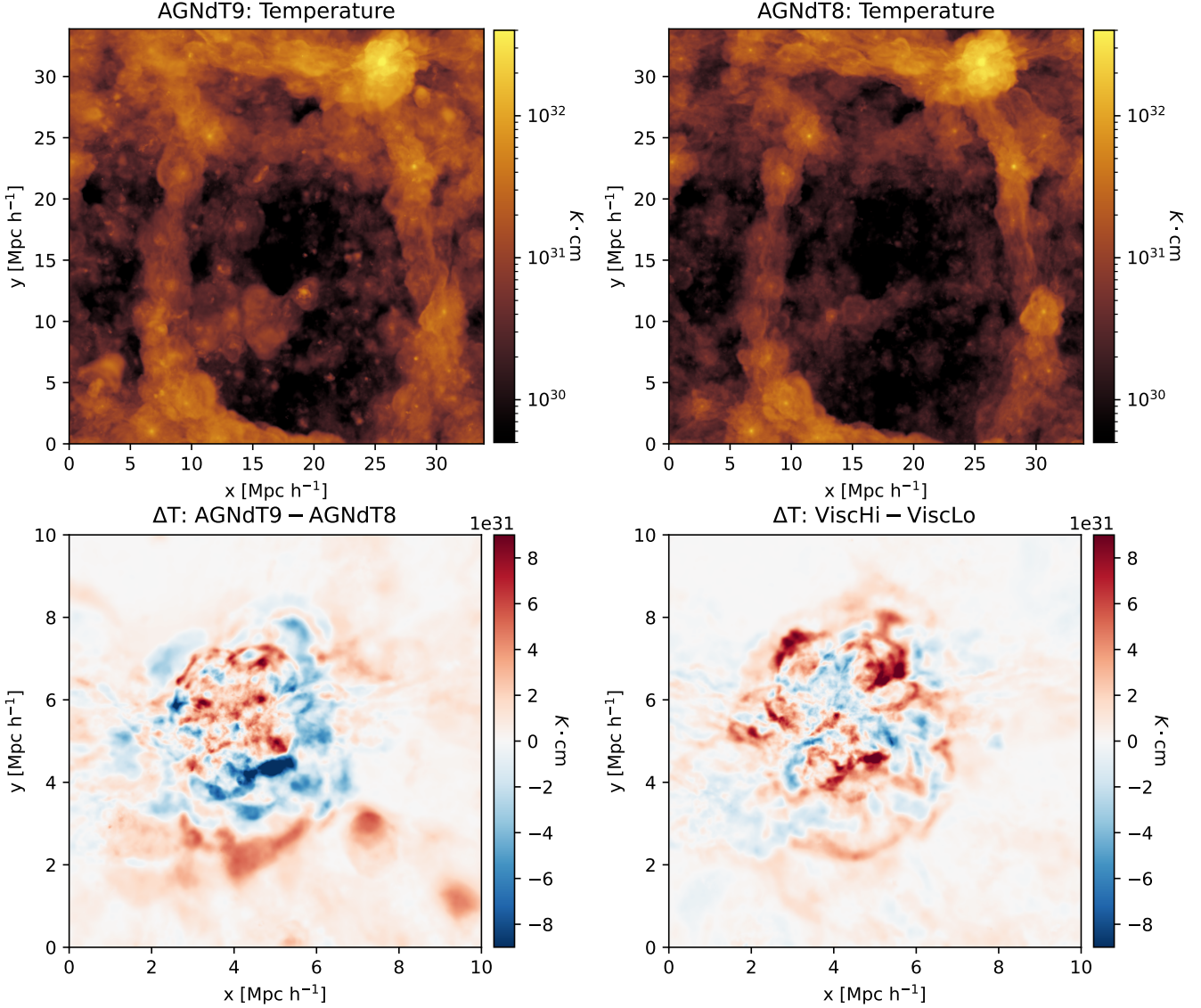


Figure 2. Temperature fields of the gaseous component across different models at $z = 0$ projected along the z -direction. The top panel shows the temperature across the entire simulation box, while the bottom panel, shows the difference in temperature profile between two models in a 10 Mpc h^{-1} region centred around the most massive halo. Gas in AGNdT9 is hotter than AGNdT8 due to higher heating temperature of BHs, while gas in ViscHi is hotter than ViscLo due to higher accretion rate which result in higher energy injection into the surrounding. The colour bar represents temperature integrated along the z -axis, and is in units of $K \cdot \text{cm}$.

the different models. We first look at the two-point statistics for different models at different epochs. The two-point correlation function (2pCF) and power spectrum ($P(k)$) are the most commonly used two-point statistics for quantifying the clustering of matter. As the Universe evolves, the strength of clustering should increase, and this remains true for all the components of the Universe we study: gas, dark matter (DM), and total matter (gas + DM + stars + BHs). This is expected from these statistics, because as the universe evolves, clustering on small scales increases as new stars form, and as gas accumulates inside galaxies (i.e. it is trapped inside dark matter potential wells). This is demonstrated in Figure A.2, where we show the evolution of $P(k)$ for the DM and total matter component.

In Figure 3, we show the $P(k)$ for AGNdT8 and ViscHi models, focusing specifically on the gas component, as we expect the

gas to be affected most significantly due to AGN feedback. Interestingly, while the general trend is for the amplitude of $P(k)$ to increase with decreasing redshift, this behaviour is not monotonic across all scales. In particular, we observe “crossover” points, i.e. where $P(k, z = z_i) < P(k, z = z_{i-1})$ for certain epochs and towards small scales. These crossover scales indicate where the clustering of gas has decreased with time. This is a signature of the onset of AGN feedback, which redistributes matter from small scales to larger scales. In Figure 3, we observe that there are crossover events at scales $k > 20 \text{ h Mpc}^{-1}$. We observe the same with the AGNdT9 and ViscHi models. It is notable that the crossover happens only for $z < 2.0$, which further indicates the time at which the dominance of AGN feedback sets in (i.e. the black holes have grown large enough that

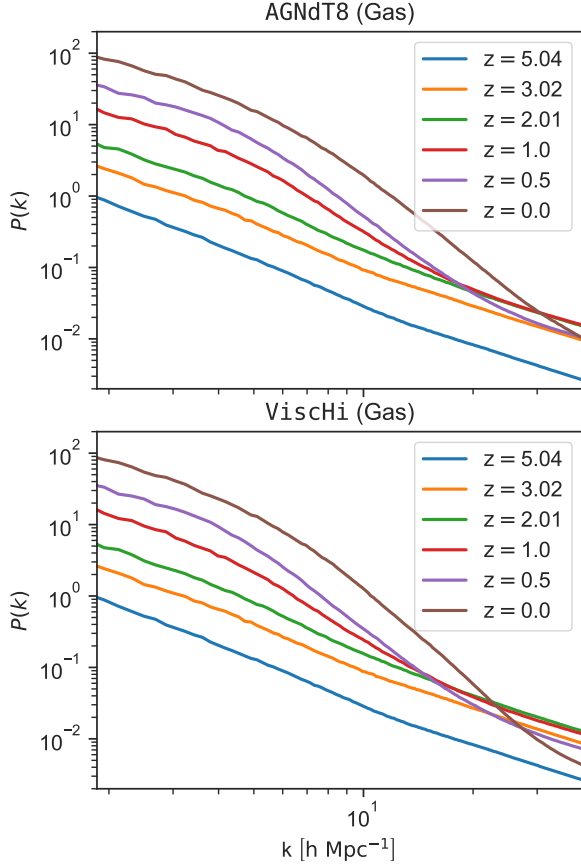


Figure 3. $P(k)$ at different epochs for AGNdt8 and ViscHi model. With evolving time, clustering increases which boosts the $P(k)$. AGN feedback re-distributes matter, which suppresses the $P(k)$. The crossover events, which are departures from general trend of $P(k)$ with z , are indicative of the onset of dominance of AGN feedback. Similar behaviour is observed with AGNdt9 and ViscLo, but with different intensity, which is studied later.

they can significantly affect the matter distribution). This motivates us to look at the $P(k)$ with a finer time resolution between $z = 1.5 - 0$.

In the two panels of Figure 4, we show the 2pCF for AGNdt8-9 and for ViscHi-Lo models at three epochs: $z = 1, 0.5, 0$. We set the lower limit of r to be $0.2 h^{-1}$ Mpc, where the finite resolution of the simulations may be dominant. We make two significant observations from these plots:

- Matter in AGNdt8 is more clustered than AGNdt9 at small and intermediate scales, but at large scales ($r > 1 h^{-1}$ Mpc), the gas is slightly more clustered in the AGNdt9 model. This is true for all three epochs. We observe the same qualitative behaviour for $z > 1$ as well.
- Matter in ViscHi is more clustered than ViscLo at small scales and almost similarly clustered at the intermediate scales.

These observations may be understood as follows:

First: between AGNdt8-9, ΔT_{AGN} is varied, which controls the efficiency and energetics of AGN feedback (refer to the Section 2.2). In AGNdt9, AGN feedback is more efficient and energetic than the feedback events in AGNdt8. Thus feedback events in AGNdt9 are better able to transport the gas from the center of the halos to the ICM, thus resulting in lower clustering at smaller scales, as feedback makes the halos less dense. The lower the radiative loss of gases within the halos, the greater is the efficiency of feedback, thus resulting in less cooling, which quenches the star formation in galaxies and further

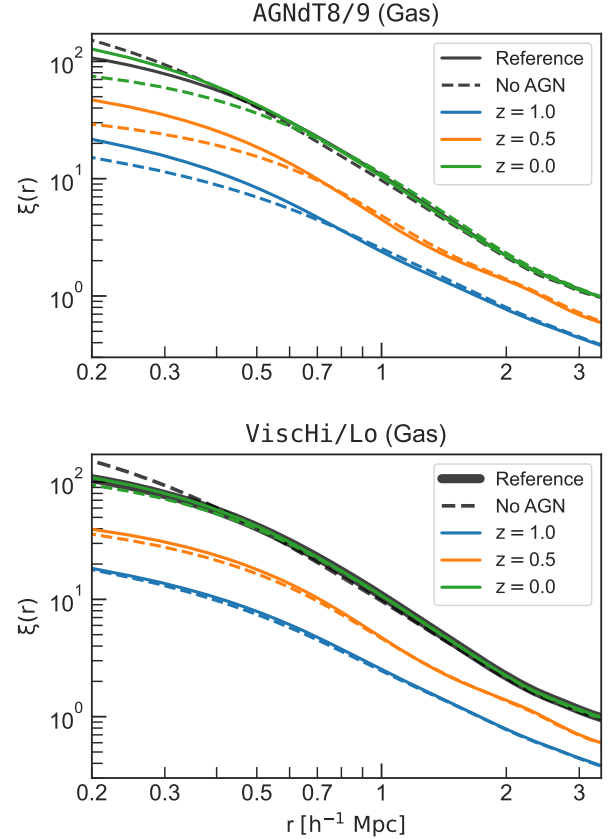


Figure 4. Comparing 2pCF of AGNdt8/9 and ViscHi/Lo. The solid lines are for AGNdt8 and ViscHi, and the dashed lines are for AGNdt9 and ViscLo, respectively. The 2pCF is boosted for AGNdt8 and ViscHi, therefore they have more clustering of matter than their counterparts. The solid black line is the 2pCF from the EAGLE reference model at $z = 0$, while black dashed line represent the 2pCF at $z = 0$ with the AGN feedback turned off.

reduces the 2pCF. The cumulative impact of these effects is to reduce the amplitude of the gas and all matter 2pCFs in the AGNdt9 compared to AGNdt8. Le Brun et al. (2014) used the cosmo-OWLS simulation suite to conduct a systematic examination of the properties of galaxy groups in response to variation of ΔT_{AGN} . They too concluded that a higher heating temperature yields more efficient AGN feedback.

At intermediate scales, there is a boost in the amplitude of 2pCF of AGNdt9, resulting in a crossover of the 2pCF of AGNdt8. This is an observational effect of AGN feedback, where the hot gas transported from the center of the galaxies and halos cools down at the intermediate scales (\sim few 100s of kpc). This causes a boost in the 2pCF at intermediate scales. BHs in AGNdt9 are more efficient than BHs in AGNdt8, and are thus able to drive more material from the center of the galaxies to ICM; also, they can drive them further than BHs in AGNdt8. This explains why the amplitude of 2pCF is larger in AGNdt9 than AGNdt8 at intermediate scales.

Second: between ViscLo and ViscHi, only the C_{visc} parameter is varied, which controls subgrid viscosity. Since AGN feedback suppresses the 2pCF at smaller scales, the suppression will be large in the case when the duration of the AGN feedback is longer. The ViscHi model has higher kinetic viscosity (i.e lower C_{visc}), which results in early onset of AGN feedback; thus we expect ViscHi to be

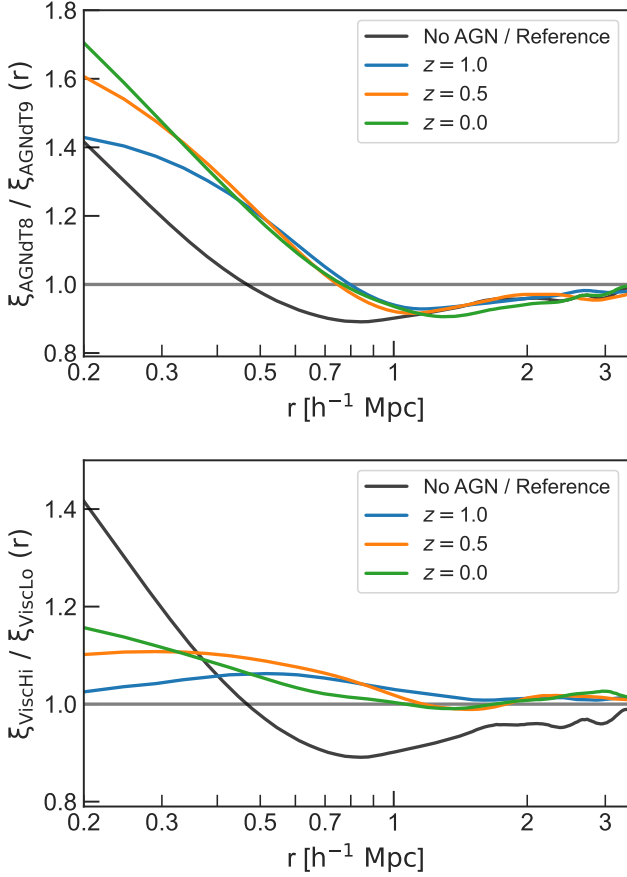


Figure 5. **Top subplot:** AGNdT9/AGNdT8 2pCF ratio shows $\sim 50\%$ more suppression at small scales and 10% boost at intermediate scales (at $z = 0$) because of more efficient AGN feedback. **Bottom subplot:** Even after early onset of AGN feedback in ViscHi 2pCF is boosted $\sim 12 - 15\%$ in ViscHi compared to ViscLo. The solid black line represents the ratio of 2pCF, The comparison of the Reference model and that without AGN is shown for $z = 0$ only.

less clustered than ViscLo. Instead, the opposite trend is observed – we return to this point later in this subsection.

In Figure 5 we show the ratios of the 2pCF for the model variations to have a better quantitative understanding of their differences. At $z = 0$, there is almost 70% stronger suppression in 2pCF in the AGNdT9 model compared to the AGNdT8 model at small scales, while at intermediate scales, there is boost of 10% in the AGNdT9 model compared to AGNdT8 model. Despite the early on-set of AGN feedback in the ViscHi model we observe a 15% boost in the 2pCF of ViscHi model compared to ViscLo model.

We also note from Figures 4 and 5 that the 2pCF in the Reference EAGLE model is bracketed by the AGNdT8–9 models, and also by the ViscHi–Lo models, although the deviation is much more prominent in the case of AGNdT8–9 models. We also note that the 2pCF from ViscHi model is very similar to the Reference EAGLE model, even after large change in the C_{Visc} parameter. This indicates that some parameters can have relatively little effect on the distribution of gas.

We noted previously that the ViscHi model exhibits a somewhat more clustered gas distribution despite the physics of the model which is expected to lead to the earlier onset of AGN feedback. To understand the reason behind this, we examine the activity of black

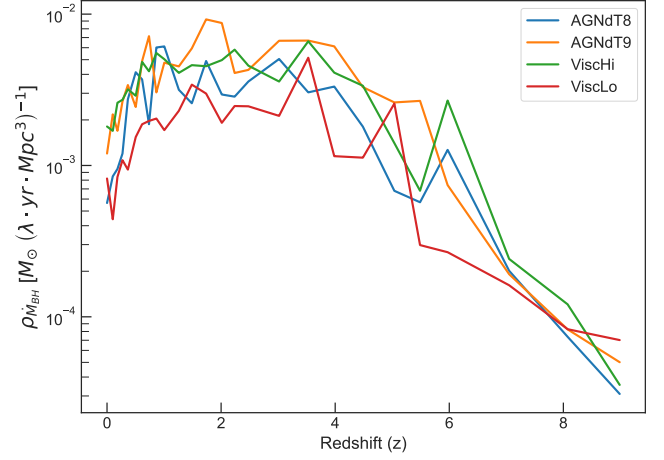


Figure 6. Black hole accretion rate density at different epochs. Higher viscosity (ViscHi) leads to higher accretion rate, thus a higher energy injection into the surrounding than ViscLo. BHs in AGNdT9 are more effective than BHs in AGNdT8, thus they have higher accretion rate so that they can grow properly. Here λ is a constant, which specifies the mass of accretion in solar masses.

holes in these models in detail. To this end, we sum the black hole accretion rate at a given epoch and divide it by the volume of the box. In other words, this corresponds to the *total* BH accretion rate density measured in the simulation.

The results are shown in Figure 6, where we can see that BH accretion rate density is higher in AGNdT9 and ViscHi models compared to their counterparts, AGNdT8 and ViscLo. BHs in AGNdT9 are more active than BHs in the AGNdT8 model; a higher accretion rate density for AGNdT9 therefore agrees with our earlier observations. A higher accretion rate density results in more redistribution of gas, and hence suppression in the 2pCF. The ViscLo model has larger C_{Visc} parameter which results in lower accretion rate of BHs (Equation 2.2) compared to BHs in the ViscHi model. Crain et al. (2015) found that having a higher subgrid viscosity i.e. lower value of C_{Visc} parameter leads to higher energy injection rate when the accretion is in viscosity-limited regime. A higher energy injection rate leads to stronger AGN feedback, and hence more redistribution of matter. This suggests the the implication of the change of accretion rate through C_{Visc} is far more complex, and the observed correlation function is due to the complex interplay of redistribution of gas through AGN feedback and the interaction of the gas with its surrounding. In this sense, it is harder to develop a direct mapping between the expected effect on the clustering of gas based on changed in subgrid viscosity than it is, for example, with changes in the AGN heating temperature.

From the above discussion, it is clear that ΔT_{AGN} , which controls the heating of particles around AGN is more influential than viscosity of gas in reshaping the matter distribution on intermediate and large scales. Thus for the rest of the analysis, we will consider just the AGNdT8 and AGNdT9 models.

4.2 Time constraints on the dominance of AGN Feedback

AGN feedback redistributes gas. The DM component remains largely unaffected by changing the AGN feedback parameters. From our previous analysis in Section 4.1 we have seen that the dominance of AGN feedback manifests as crossover points in the power spectrum

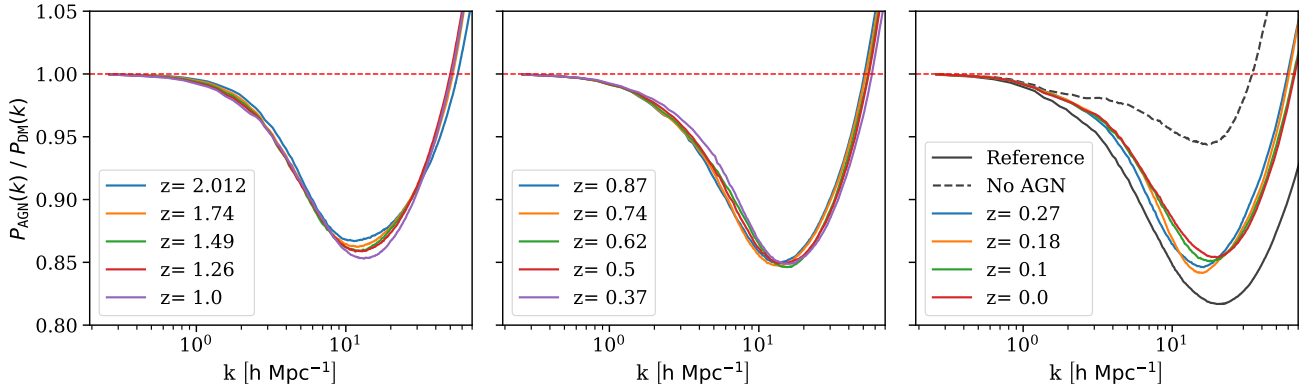


Figure 7. Ratio of the total matter and the corresponding DM power spectrum at different epochs for AGNdT8 model. The solid black curve shows the same ratio for the EAGLE Reference model at $z = 0$ while the dashed black line shows the same ratio with AGN feedback turned off at $z = 0$.

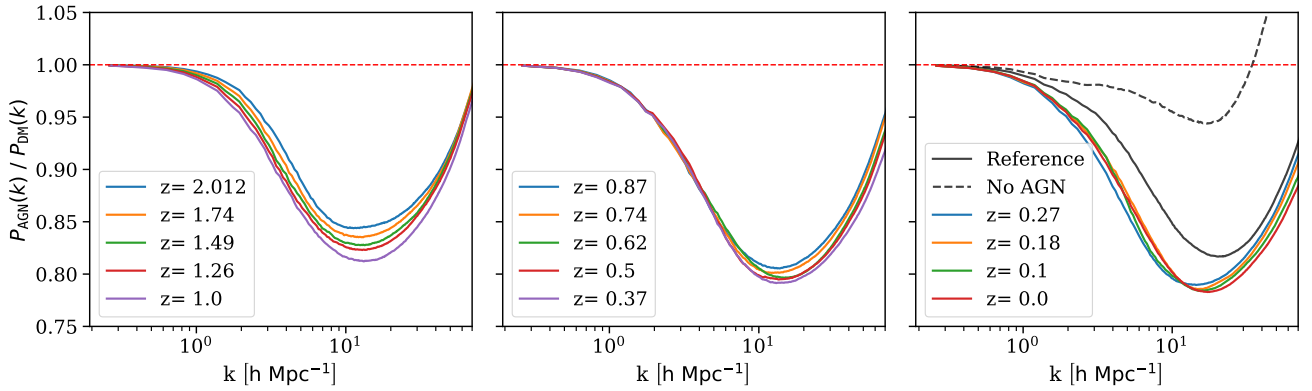


Figure 8. Ratio of the total matter and the corresponding DM power spectrum at different epochs for AGNdT9 model. The solid black curve shows the same ratio for the EAGLE Reference model at $z = 0$ while the dashed black line shows the same ratio with AGN feedback turned off at $z = 0$.

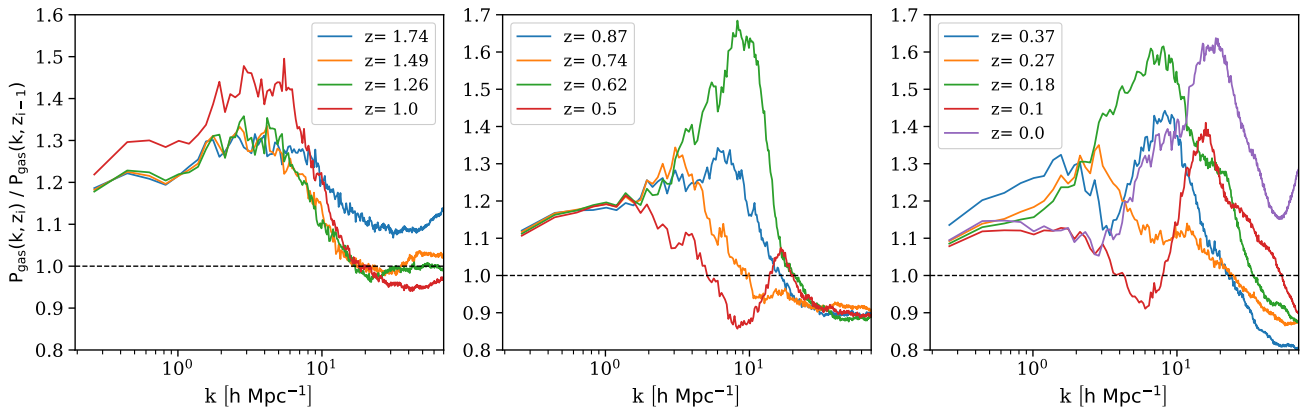


Figure 9. Ratio of power spectrum of gas between two consecutive epochs for AGNdT8 model. The panels are arranged in decreasing redshift if we go from left to right. The ratio significantly falls below 1 at $z = 1$ (first panel from left) which marks the onset of dominance of AGN feedback.

on small scales. Thus, to put further constraints on the effects of AGN feedback, we first look at the ratio of $P(k)$ between the total matter and DM at different epochs using finer snapshot spacing between consecutive epochs. This enables us to understand the temporal and spatial onset of AGN feedback in the EAGLE model.

From Figures 7 and 8, we observe the same features that we observed in Figure 4, i.e., there is stronger suppression of $P(k)$ in

AGNdT9 compared to AGNdT8, which is explained in the discussion following Figure 4. Additionally, we also observe that suppression between each consecutive epoch is more in the case of AGNdT9 than AGNdT8 model, which is due to the higher efficiency of AGN feedback in AGNdT9. We also note that the baryonic effects are visible on scales as large as 0.8 h Mpc^{-1} even at $z \approx 2$. In Figures 7 and 8 we also plot the same ratios for the EAGLE Reference model (solid

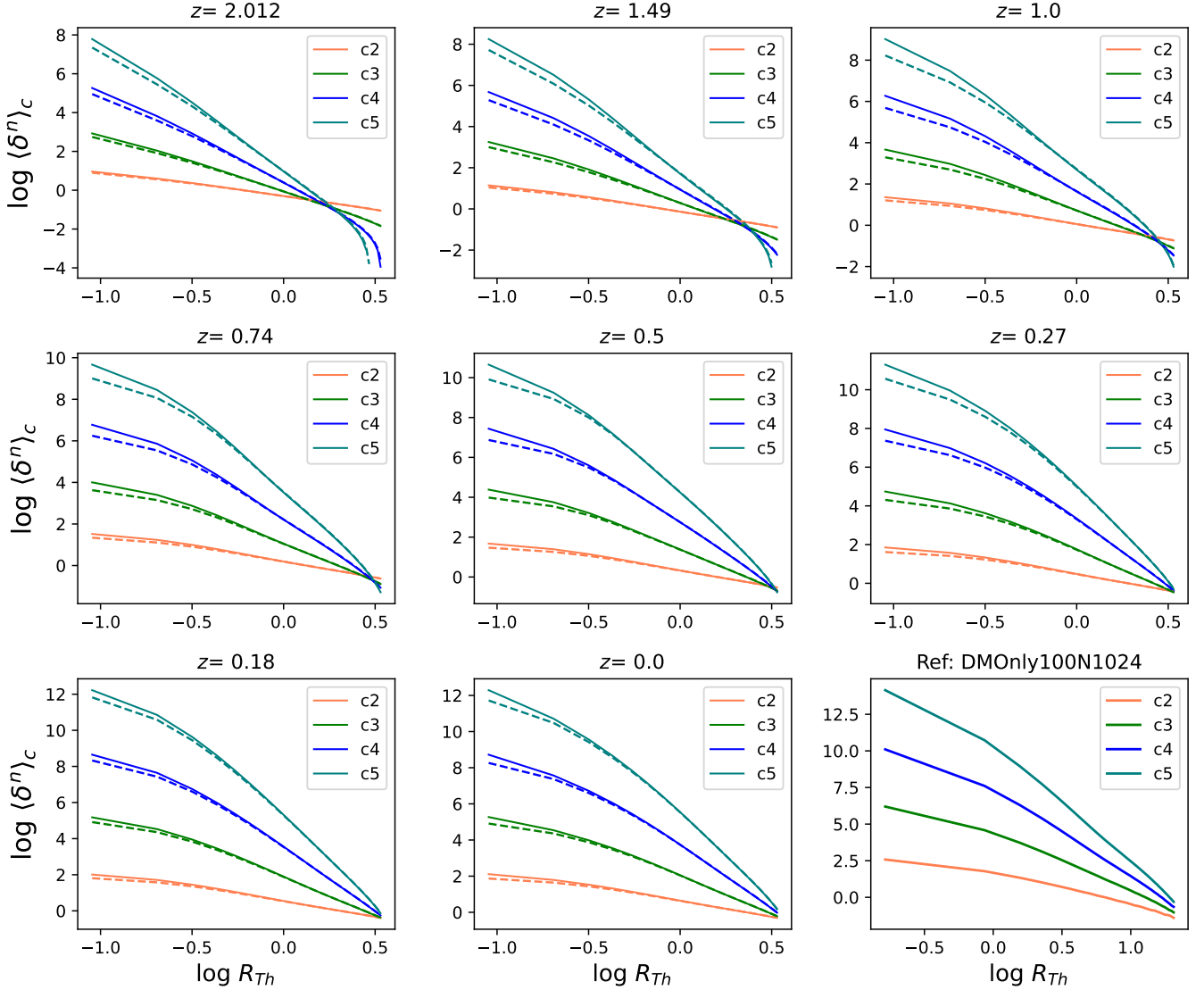


Figure 10. Cumulants of the gas density field at different epochs for AGNdT8 and ANGDt9. The solid lines represent the cumulants for AGNdT8 model and the dashed lines show the cumulants of the AGNdT9 model. The labels $c\{i\}$ represent the i -th cumulant. The bottom-right panel, shows the cumulants from the DM-only simulation of the Eagle project for reference with box size of 100 Mpc and 1024 particles along each side. BHs in AGNdT9 redistributes gas more efficiently than those in AGNdT8, which results in more homogeneous distribution of gas observed as reduced values for cumulants on small scales.

black curve) and a model variation with AGN Feedback turned off (dashed black curve), both at $z = 0$. We again note that the EAGLE Reference model is bracketed by the AGNdT8–9 models. Moreover, even turning off the AGN feedback results in a change in the $P(k)$ by $\approx 5\%$ at the intermediate scales ($\approx 10 \text{ h Mpc}^{-1}$). Interestingly, the suppression in $P(k)$ for all models is most pronounced on scales $k \approx 15\text{--}20 \text{ h Mpc}^{-1}$, with the maximal suppression (relative the dark matter-only case) being of the order of 20–25%. These observations further demonstrate the need to better understand AGN feedback to accurately model the matter distribution at scales, $k > 1 \text{ h Mpc}^{-1}$ for precision cosmology applications.

Finally, we look at the crossover events to determine when AGN feedback started to dominate. We plot the ratios of the power spectrum of gaseous component between two consecutive epochs and check when the ratio significantly falls below 1. The result is presented in Figure 9 for AGNdT8. We can see that the ratio significantly

falls below one at $z = 1$. As we go to lower redshifts, we see that the decrease of the ratio becomes sharper, and we start to notice the fall at intermediate scales as well. The suppression at intermediate scales appears first at $z = 0.74$ and is prominent by $z = 0.5$. Thereafter, the effects of AGN feedback start to equilibrate, and for $z < 0.5$, we notice the magnitude of slope decreases, and for some epochs, we do not see the ratio falling below one. These observations suggest that, in the EAGLE model, AGN feedback starts to dominate at $z \approx 1 - 0.7$, and after that, effects start to equilibrate with the surrounding. For the AGNdT9 model, we observe that AGN feedback starts to dominate somewhat earlier, $z = 1.49 - 1$, due to increased efficiency of BHs. McAlpine et al. (2017) Found similar results by comparing the star formation rate (SFR) and black hole accretion rate in EAGLE Reference Model. Since AGN feedback, quenches the star formation in galaxies, SFR can be used as an indirect tracer for AGN feedback. They found a rapid decline in the SFR in halos with

$M_{200} \approx 10^{12} M_{\odot}$ which has a median redshift, $z = 1.9$, which is very close to our inference from the evolving $P(k)$.

4.3 Cumulants of the matter distribution

As described in Section 3.2, we use cumulants of the matter distribution to probe the non-linear evolution of the matter density field. We compute the cumulants at different epochs. We first compute the density field for the three components of the matter. Then we smooth the density field with a Gaussian kernel and compute the cumulants. We repeat the step for different values of the smoothing scale, R_{Th} , and at different redshifts. For our analysis, we use $\frac{2}{752} \text{BoxSize} < R_{Th} < 0.1 \text{BoxSize}$ where BoxSize is the size of the simulation box and is equal to $33.885 h^{-1} \text{Mpc}$. The lower bound is set using the Nyquist frequency, while the upper bound is set empirically. We plot the cumulants of the gas at different epochs for AGNdT8 and AGNdT9 in Figure 10. In the bottom-right panel, we plot the cumulants from the DM-only simulation of the EAGLE project for reference.

Cumulants are an essential tool to detect deviations from Gaussianity. A Gaussian field is characterised by only the first two moments or cumulants, and higher-order cumulants are zero; non-zero higher-order cumulants therefore indicate non-Gaussian behavior. It is therefore expected that when smoothed over small scales, the density field is more non-Gaussian than when smoothed with a larger filter. Indeed, it is expected that for large enough values of the smoothing scale, all the higher-order cumulants should converge to zero, and only the variance (i.e., the second-order cumulant) should survive. Also, as the density field evolves, the irregularities grow stronger and stronger; therefore, we expect that as z decreases, the amplitude of the cumulants increases.

Looking at the panels of Figure 10, we observe that the variance (i.e second order cumulant) is enhanced in AGNdT8 compared to the AGNdT9 model for all smoothing scales. This is because BHs in the AGNdT9 model are more efficient than BHs in the AGNdT8 model, hence in the AGNdT9 model, gas particles are distributed over larger distances in the simulation box, thereby making the gas distribution is more uniform in AGNdT9 compared to the AGNdT8 model. We also observe a significant difference between the variance of the two models only for $z \leq 1$, which agrees with our previous constraints on when the AGN feedback starts to dominate. We also observe that, at $z = 0$, the cumulants of AGNdT8/9 do not vary much with respect to the DM-only case.

We observe similar behaviour in the higher order cumulants as well, i.e., cumulants of AGNdT8 are enhanced compared to cumulants in AGNdT9 – shown as the different coloured lines in Figure 10. Again, this is a manifestation of enhanced clumpiness in the gas distribution in AGNdT8, as higher-order moments are predominantly affected only by higher density regions of the density field ($\delta \gg 1$).

We also note that the cumulants of gas in the full hydrodynamic model at $z = 0$ are very similar to the cumulants of DM in the DM-only simulation with larger box size at $z = 0$. In particular, the cumulants in the hydrodynamic case have simply been shifted to the right of those in the DM-only simulation. Baryonic processes in the hydrodynamic model act to redistribute the gas and thus effectively reduce the irregularities which result in smaller magnitude of the cumulants at smaller smoothing scales. Again, this observation further stresses the importance to understand the baryonic processes.

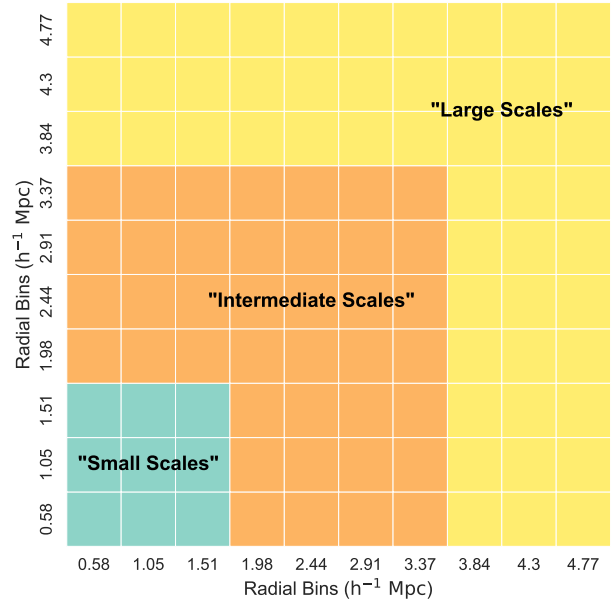


Figure 11. An illustrations depicting different scales in the 3pCF matrix.

4.4 The three-point correlation function

Compared to the 2pCF, the three-point correlation function (3pCF) is relatively understudied as a tool for quantifying the matter distribution in hydrodynamic simulations. This is largely due to the increased computational cost ($\approx O(n^3)$ given n tracers). However, the 3pCF may contain more information on the distribution of matter than what is simply contained in the 2pCF. This section discusses our result from the 3pCF analysis of AGNdT8, AGNdT9.

Our observations from Figure 10 already give us an initial impression that the non-Gaussian distribution of the gas changes if we vary the AGN feedback, as we observe that the amplitude of the higher order cumulants ($n > 2$) are significantly different at small smoothing scales between AGNdT8 and AGNdT9. However, it does not tell us anything about the strength of the changes because cumulants do not consider the correlation between points. This motivates the use of the 3pCF. In order to compute this from our simulation data, we first take 20 random uniform samples from the entire dataset containing 0.1% of the total gas particles; this is done to reduce computational cost. This step does not affect our conclusions; we have checked explicitly that our results are converged with respect the size of our sub-sample (Appendix C). We then compute the 3pCF on each of these datasets, and then we get the final 3pCF matrix by taking the average of the 3pCFs obtained from the subsets. We consider triangles with sides up to 5 Mpc only, and we bin the triangles into ten bins per side. We then estimate the error in the 3pCF value for a given radial bin by taking the standard deviation between the 3pCF for the same radial bin from the samples.

In Figure 11, we present an illustration of the 3pCF matrix, highlighting the different scales where the information content is stored. The x and y ticks are the average of the upper and lower bound of the radial bin, so in given bin (r_1, r_2) , we consider triangles whose sides length $(r_1 \pm \Delta r, r_2 \pm \Delta r)$ Mpc, where $\Delta r = 0.25$ Mpc. We loosely define “small scales” as the regime where galaxies may dominate i.e triangles with side lengths up to 1.7 Mpc. “Intermediate scales” denote regions of a few virial radii of the most massive haloes in the simulation i.e., galaxy clusters with side lengths upto 3.6 Mpc. Finally, “large scales” are defined as anything beyond this scale.

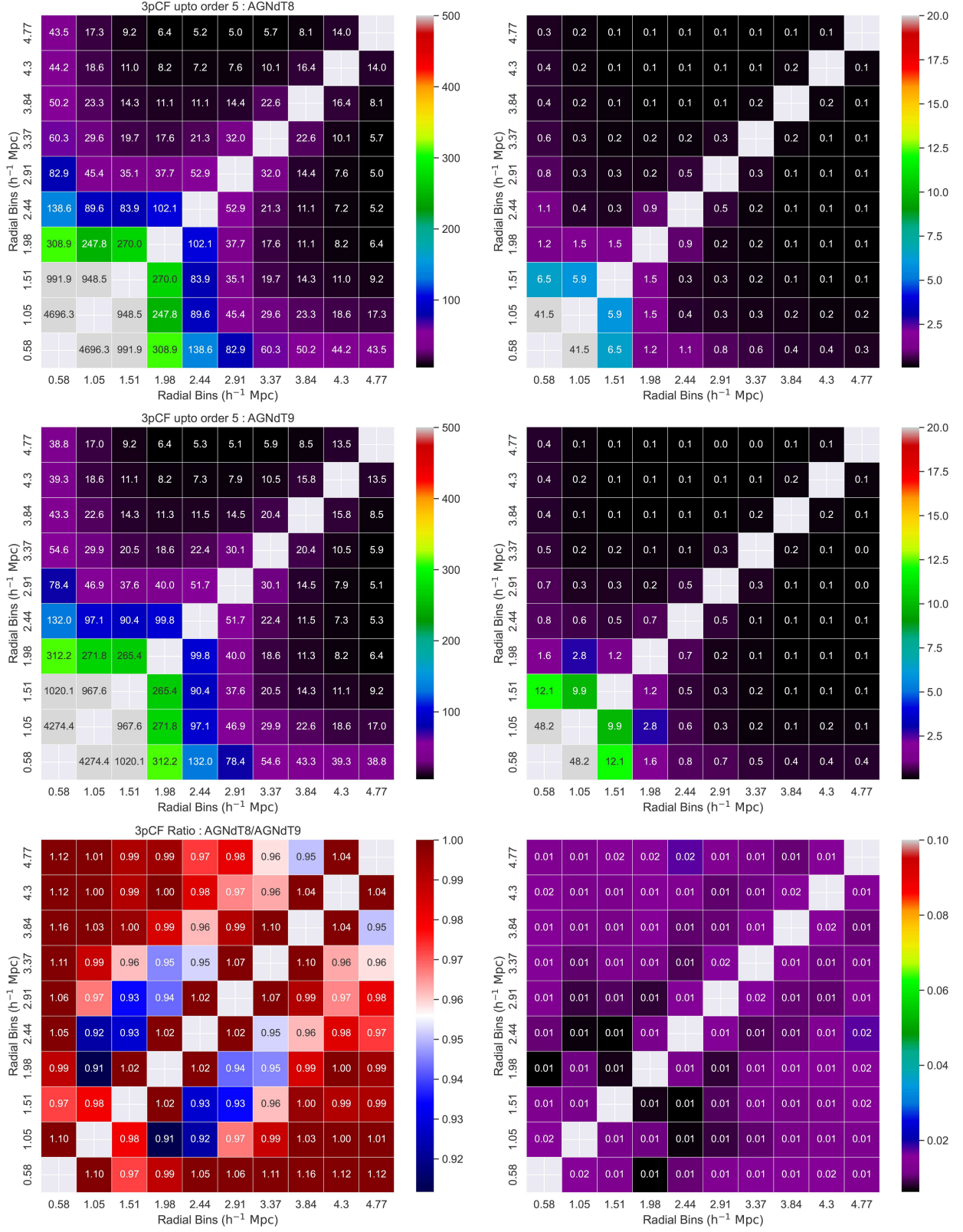


Figure 12. 3pCF of AGNdT8 (first row), AGNdT9 (second row) and their ratio (third row). The right-hand column shows the error in measurement for the corresponding bin. The 3pCF follows similar behaviour to 2pCF. It decays rapidly from small scales to large scales. At small scales, the DDD count is significantly higher than RRR count, as matter is very clustered inside galaxies. As we move to large scale, matter distribution becomes more homogeneous and the DDD count falls and becomes more or less constant. From the third row we see, varying AGN feedback does not significantly change the 3pCF. More gas is transported from the centre of haloes to the intermediate scales when the AGN feedback is more efficient, which results in a higher *DDD* count, boosting the 3pCF on these scales.

We caution that these demarcations are determined qualitatively, and there is no universally accepted limit on the choice of different scales.

In Figure 12, we present the 3pCF for AGNdT8 and AGNdT9. We consider the $\ell_{max} = 5$, which corresponds to the maximum value of ℓ in Equation 3.7. The top panel presents the 3pCF and error for the AGNdT8 model, the middle panel presents the 3pCF and error for the AGNdT9 model, and the bottom-most panel compares the ratio of the 3pCF between these two models. There are two major observations: first, the 3pCF has a larger amplitude on small scales when the sides of the triangle are less than $1.5 h^{-1}$ Mpc, which is the scale of the galaxies and the interaction between galaxies. Beyond this scale, the 3pCF decays rapidly. Second, on large scales, i.e., for triangles with side lengths greater than $3.5 h^{-1}$ Mpc, the 3pCF becomes more or less constant and does not decay rapidly. These observations hold true for both the AGNdT8 and AGNdT9 models.

On small scales, we expect the amplitude of the 3pCF to be higher because the matter distribution is very clumpy containing, for example, star-forming regions and gas inside galaxies. It is more likely to find gas particles in triangular configurations of small side lengths as we approach the scales of the galaxies; this is observed as a boost in the 3pCF. On larger scales, the Universe becomes more homogeneous. Beyond a certain length scale, this decrease in DDD relative to RRR starts to saturate, and we do not observe much change in the 3pCF. The above observation, in general, is valid for 2pCF as well.

From the bottom-most panel of Figure 12, which compares the 3pCF of AGNdT8 and AGNdT9, we see that 3pCF of AGNdT8 and AGNdT9 have similar magnitudes. 3pCF of AGNdT9 is slightly higher at the intermediate scales. We argued in Section 4.1 that BHs in AGNdT9 are more efficient than BHs in AGNdT8, i.e they have less radiative loss, which results in more effective gas transport from the center of the halos to intermediate scales (i.e., CGM), which results in higher DDD count at intermediate scales in the AGNdT9 model. The same argument holds for large scales as well; however, at large scales, this change is very negligible, so the ratios for most of the triangle configuration remain ≈ 1 .

For completeness, we also checked the 3pCF from the `ViSchi-Lo` model and from these model also, we reach the same conclusion. However, the difference in the magnitude between the two models is very less ($\approx 2\%$ on average in a given radial bin) compared AGNdT8–9, hence we don not show that here.

5 DISCUSSION

We found that the effects of varying the kinematic viscosity have very tiny effect on the 2pCF or $P(k)$. Additionally, directly mapping the effects of the viscosity parameter with the resulting impact on the gas through feedback is rather difficult. Using EAGLE, we can only study the effects upto $3.3 h^{-1}$ Mpc ≈ 5 Mpc. It might be the case that effect of viscosity manifest on scales larger than this. Therefore to understand the effect of viscosity we need simulations with larger box size and with higher resolution, with much more drastic variations in the viscosity. Moreover, to study the large-scale effects of changing these parameters, we need a much bigger simulation box size, at least of the orders of 100 Mpc, that would enable us to measure the effect on the 2pCF on the scale of 10s of Mpc. Indeed, the subgrid treatment of gas viscosity used in this work may be too crude, and fails to capture complex kinematics of gas flows around black holes.

The same caveat holds for the cumulants as well - in principle at larger smoothing scales, the higher order cumulants should converge which we also observe in Figure 10. However it may be the case that cumulants measured in simulations with different AGN feed-

back prescriptions converge on scales larger than those probed in the simulations used here. That being said, however, the *qualitative* comparison for AGNdT8 and AGNdT9 remain valid. [Borrow et al. \(2022\)](#) showed that globally averaged properties in a full cosmological volume differ between clone simulations, but the deviation diminishes for box size > 25 Mpc; this is because of stochasticity in the subgrid models underlying such simulations. For more robust constraints on measurements involving two-point statistics, we need multiple runs with the same configuration (i.e., cosmological and galaxy formation models) to enable better statistical fidelity in our estimates of the cumulants.

It may also be the case that the size and resolution of the simulations we use are not enough to take full advantage of the 3pCF. In principle, the three-point function should contain more information about changes to the large-scale matter distribution due to the effects of feedback than just $P(k)$ or the 2pCF, but this is not immediately apparent based on our observations in Section 4.4. The 3pCF estimator is more likely to be affected by our limited sample size than the 2pCF and it will therefore be interesting to revisit this exploration with larger volume hydrodynamical simulations, particularly those that also vary subgrid parameters relating to feedback more widely than the set considered in this work.

Nevertheless, in this work we have noted how changing the strength of AGN feedback is able to introduce change in the two-point statistics and higher-order cumulants by several tens of percentage even at scales $k \approx 1 h \text{ Mpc}^{-1}$. Given that upcoming surveys are hoping to constrain cosmological parameters using observables measured on scales $0.1 \leq k \leq 10 h \text{ Mpc}^{-1}$ to 1% or better, this further strengthens the case for why we need a better understanding of the baryonic physics and feedback effects.

6 CONCLUSIONS

This work investigates the effects of AGN feedback on the large-scale matter distribution in cosmological, hydrodynamical simulations focussing on the redistribution of gas in and around dark matter haloes. In particular, we use variations in the subgrid physics model of the EAGLE simulations ([Schaye et al. 2015](#)) to study the impact of varying the BH feedback model on two- and three-point statistics used to the characterise the matter distribution. Our main findings are summarised below:

(i) We find that the efficiency of AGN feedback (controlled by the parameter ΔT_{AGN} , which determines how much the surrounding particles are heated) is crucial in deciding the gas distribution on large scales. More efficient AGN feedback results in a stronger suppression of the two-point correlation function (2pCF) and the power spectrum, $P(k)$, at small scales ($r < 1 h^{-1}$ Mpc) and enhances it at intermediate scales ($1 h^{-1} \text{ Mpc} < r < 10 h^{-1} \text{ Mpc}$, Figure 4). This is because, more efficient AGN feedback leads to more gas transport from the galaxies' center to the circumgalactic medium (CGM).

(ii) Increasing the viscosity of gas while keeping the efficiency of AGN feedback fixed results in earlier onset of AGN feedback. Thus, one would naively expect to see the redistribution of gas from the centre of galaxies to the CGM for the model with higher viscosity and thus a reduced clustering at small scales. However, we instead observe the reverse effect – that the higher viscosity model is more clustered than the low viscosity model (Figure 4). This warrants further investigation, potentially involving larger simulations than the ones we have used in this work.

(iii) By considering the ratio of $P(k)$ for gas between two consecutive epochs, we are able to narrow down the redshift range during

which effects of AGN feedback were most dominant as $z = 1 - 0.74$ (Figure 9). However, this time scale also varies depending on the efficiency of the BHs – more efficient BHs i.e. in AGNdT9 results in the effects showing up earlier at $z = 1.49 - 1$ (Figure A.3).

(iv) Varying AGN feedback does not change the cumulants of dark matter distribution because AGN feedback mainly affects the gas, and the any net effect on the DM is negligible. More efficient AGN feedback results in a less clumpy gas distribution, resulting in a reduction in the magnitude of the cumulants compared to the model with less efficient AGN feedback (Figure 10).

(v) The 3pCF of the gas distribution (Figure 12) also qualitatively exhibits similar behavior to the 2pCF. With large amplitude on small scales ($r < 1.7 \text{ h}^{-1} \text{ Mpc}$), it decays rapidly at intermediate scales ($1.7 < r < 3.6 \text{ h}^{-1} \text{ Mpc}$) and becomes more or less constant beyond this ($r > 3.6 \text{ h}^{-1} \text{ Mpc}$), which implies that the value of the 3pCF changes negligibly with increasing side length of the triangles i.e. approaching homogeneity in large scale.

(vi) Varying the efficiency of AGN feedback does not significantly change the 3pCF. The only noticeable changes are on intermediate scales, as more gas is transported from the centre of the haloes to the intermediate scales when the AGN feedback is more efficient. This boosts the 3pCF at intermediate scales. These observations are consistent with what we concluded with the 2pCF. The present simulations may, however, be too limited in the size/resolution, to extract the full information content in the 3pCF.

The present work adds to the growing body of work demonstrating the importance of considering the effects of galaxy formation and feedback on the large-scale matter distribution, particularly given the ambitions of precision cosmology. We have shown how relatively small variations in parameters that are, in general, poorly constrained can leave imprints on the matter distribution from anywhere between 5-25%, depending on the model, redshift, and the scales of interest. Our work also shows some of the limitations of finite box size and the scope of model variations we have considered; in particular, it would be illuminating to consider the use of the 3pCF in characterising the gas distribution in larger simulations that also incorporate a more wide range of feedback mechanisms. New generations of hydrodynamical simulations like the FLAMINGO project (Schaye et al. 2023) provide the perfect opportunity to pursue these scientific questions.

ACKNOWLEDGEMENTS

SB is supported by the UK Research and Innovation (UKRI) Future Leaders Fellowship [grant number MR/V023381/1]. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1 and ST/R002371/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

DATA AVAILABILITY

The data supporting the plots within this article are available on request to the corresponding author and can also be access from the following repository.

REFERENCES

- Ade P. A. R., et al., 2016, *A&A*, 594, A13
 Bernardeau F., 1992, *ApJ*, 392, 1
 Bondi H., Hoyle F., 1944, *Monthly Notices of the Royal Astronomical Society*, 104, 273
 Borrow J., Schaller M., Bahe Y. M., Schaye J., Ludlow A. D., Ploekinger S., Nobels F. S. J., Altamura E., 2022, *arXiv e-prints*,
 Chisari N. E., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 480, 3962
 Crain R. A., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1937
 Fry J. N., 1984, *ApJ*, 279, 499
 Fry J. N., 1985, *ApJ*, 289, 10
 Hearin A. P., Zentner A. R., Ma Z., 2012, *Journal of Cosmology and Astroparticle Physics*, 2012, 034
 Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, *Monthly Notices of the Royal Astronomical Society: Letters*, 461, L11
 Huff E. M., Hirata C. M., Mandelbaum R., Schlegel D., Seljak U., Lupton R. H., 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 1296
 Huterer D., Takada M., 2005, *Astroparticle Physics*, 23, 369
 Ivezić Z., et al., 2019, *The Astrophysical Journal*, 873, 111
 Laureijs R., et al., 2011, *arXiv preprint arXiv:1110.3193*
 Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 1270
 McAlpine S., Bower R. G., Harrison C. M., Crain R. A., Schaller M., Schaye J., Theuns T., 2017, *MNRAS*, 468, 3395
 Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 1958
 Mo H., Van den Bosch F., White S., 2010, *Galaxy formation and evolution*. Cambridge University Press
 Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton university press
 Philcox O. H. E., Slepian Z., Hou J., Warner C., Cahn R. N., Eisenstein D. J., 2021, *Monthly Notices of the Royal Astronomical Society*, 509, 2457
 Pillepich A., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 473, 4077
 Rosas-Guevara Y. M., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 1038
 Schaye J., et al., 2010, *Monthly Notices of the Royal Astronomical Society*, 402, 1536
 Schaye J., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 521
 Schaye J., et al., 2023, *The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys (arXiv:2306.04024)*
 Schneider A., Teyssier R., 2015, *Journal of Cosmology and Astroparticle Physics*, 2015, 049
 Slepian Z., Eisenstein D. J., 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 4142
 Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 1105
 Springel V., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 475, 676
 Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *The Astrophysical Journal*, 761, 152
 Troxel M. A., et al., 2018, *Physical Review D*, 98, 043528
 Villaescusa-Navarro F., 2018, Pylians: Python libraries for the analysis of numerical simulations, *Astrophysics Source Code Library*, record ascl:1811.008 (ascl:1811.008)
 Vogelsberger M., et al., 2014, *Nature*, 509, 177
 de Jong J. T., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, *Experimental Astronomy*, 35, 25
 van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 3649

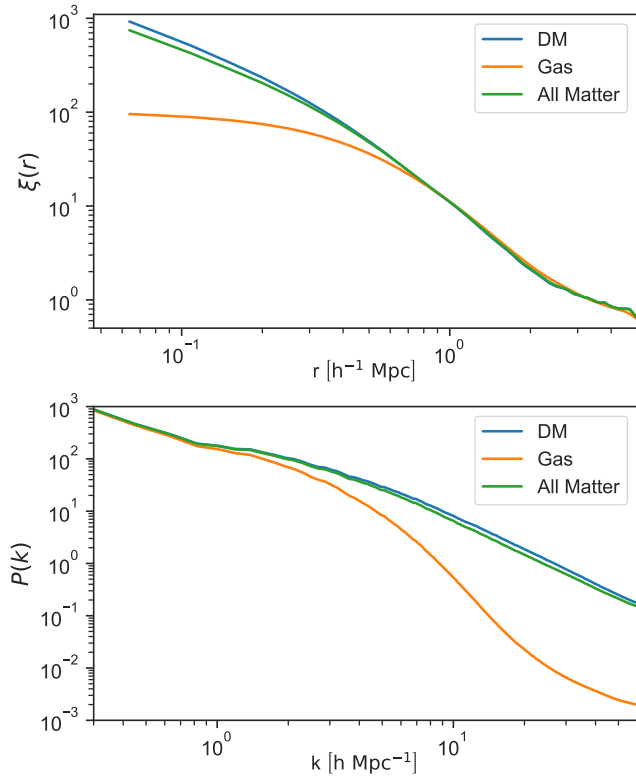


Figure A.1. 2pCF and $P(k)$ for AGNdT9 model. This figure compares the 2-point estimate for different components of the matter. The DM component remains invariant in all model variations. Only the gas and the total matter (DM + gas + stars + BHs) components get affected. We are interested in the change of gas distribution as that is the component that is most significantly affected by variations in the AGN feedback model.

APPENDIX A: TWO POINT ESTIMATES

In this section, we take a closer look at the 2-point estimates. Figure A.1 compares the 2pCF and $P(k)$ for different components of the matter distribution in the AGNdT9 model. We observe similar behavior in all the other models. The gaseous components of the matter distribution is affected the most due to AGN feedback, as AGN feedback moves gas from the centre of the haloes and redistributes it to the exterior. This also quenches star formation in galaxies, further lowering the clustering at small scales, seen as a suppression of the 2pCF and $P(k)$ on these scales. However, we do not observe the same amount of suppression in the total matter component that we observe in the gaseous component because a significant contribution to the total matter comes from the DM, which remains largely unaffected due to AGN feedback.

For completeness, in Figure A.2, we also present the $P(k)$ of all the models we studied at different epochs. We observe crossover events discussed in Section 4.1 in the gaseous component in all models at $z \approx 1 - 1.5$, which agrees with our constraints on the dominance of AGN feedback from AGNdT8 and AGNdT9 models.

To determine the time scale at which AGN feedback starts to dominate, we plot in Figure A.3 the ratio of $P(k)$ between two consecutive epochs for AGNdT9, similar to the Figure 10. We observe that, in this case, dominance of AGN feedback starts a bit earlier, at around $z = 1.49 - 1.26$. This is due to the higher ΔT_{AGN} parameter for the black holes in AGNdT9 model that make the black holes more

efficient in heating the surrounding and transportation of the gas due to AGN feedback. Also compared to the same plot for AGNdT8 in Figure 10, the changes in AGNdT9 are much more drastic, particularly at late times, i.e $z = 0.18 - 0$. We observe a stronger and rapid change in characteristics of the curve (i.e change in the slopes of the curve) in the AGNdT9 model.

APPENDIX B: EFFECT OF BOX SIZE

Cumulants are highly sensitive to the size and resolution of the box. To have a complete large-scale picture, we need a larger simulation box size without compromising resolution. To demonstrate how different box size and resolution can effect the statistical quantities, we plot the distribution of the first cumulant, i.e the PDF of the density field for different smoothing scales for simulations with different box sizes.

In Figure B.1, we plot the distribution of gas at various smoothing scales for AGNdT8, in Figure B.2 we plot the same quantities for the EAGLE reference model with a box size of 100 Mpc. Furthermore, in Figure B.3, we compare these distributions to the IllustrisTNG-300-1 simulation. The IllustrisTNG suite (Pillepich et al. 2017) is comprised of multiple runs with different box sizes and resolutions. We use the TNG300 – 1 run which has a periodic box size of $L = 205 h^{-1} \text{ Mpc} \approx 300 \text{ Mpc}$ on a side and uses 2×2500^3 resolution elements, and uses the PLANCK 2016 (Ade et al. 2016) cosmology.

We see that the box size affects the distribution because the distribution in AGNdT8 is a skewed Gaussian, whereas, for TNG300-1, it becomes a narrower and peaks around 0 as we increase the smoothing scale. We also note that the differences between the AGNdT8 model and the EAGLE reference model is not as significant as the differences between the AGNdT8 and the TNG300-1 model. This further strengthens the need to larger box sizes as the statistical properties can be significantly different for the larger box. This variation will be more prominent if we look at the higher order cumulants from TNG300-1. The larger box size in TNG300-1 results in a more uniform distribution, which will also lower the magnitude of the cumulants. However, we note that the qualitative comparison between AGNdT8, AGNdT9 made throughout Section 4 remains valid.

APPENDIX C: EFFECT OF SAMPLING ON 3PCF

In Section 4.4 we computed the 3pCF for the distribution of gas for AGNdT8 and AGNdT9. For computing the 3pCF, we first take 20 random uniform samples from the entire dataset containing 0.1% of the total gas particles; this is done to reduce computational cost. In this appendix, we consider the error induced by sub-sampling the gas distribution.

In Figure C.1 we compute the ratio of 3pCF of AGNdT9 with 0.1% subset of the data to the 3pCF of AGNdT9 with 0.05% sample of the complete dataset. Similarly in Figure C.2 we compute the ratio of 3pCF of AGNdT9 with 0.1% subset of the data to the 0.5% sample of the complete dataset, and in C.3 we compute the ratio of 3pCF of AGNdT9 with 0.5% subset of the data to the 0.05% sample of the complete dataset. From these figure we can see that our choice of sampling the data doesn't affect the results significantly and the small deviation (on average 0.3%) is just due to random sampling.

The only effect of sampling is that, by taking a small subset of the data we are unable to retain the minute details of the matter distribution. However this does not significantly limit our understanding of matter distribution at the large scale. The large scale structures

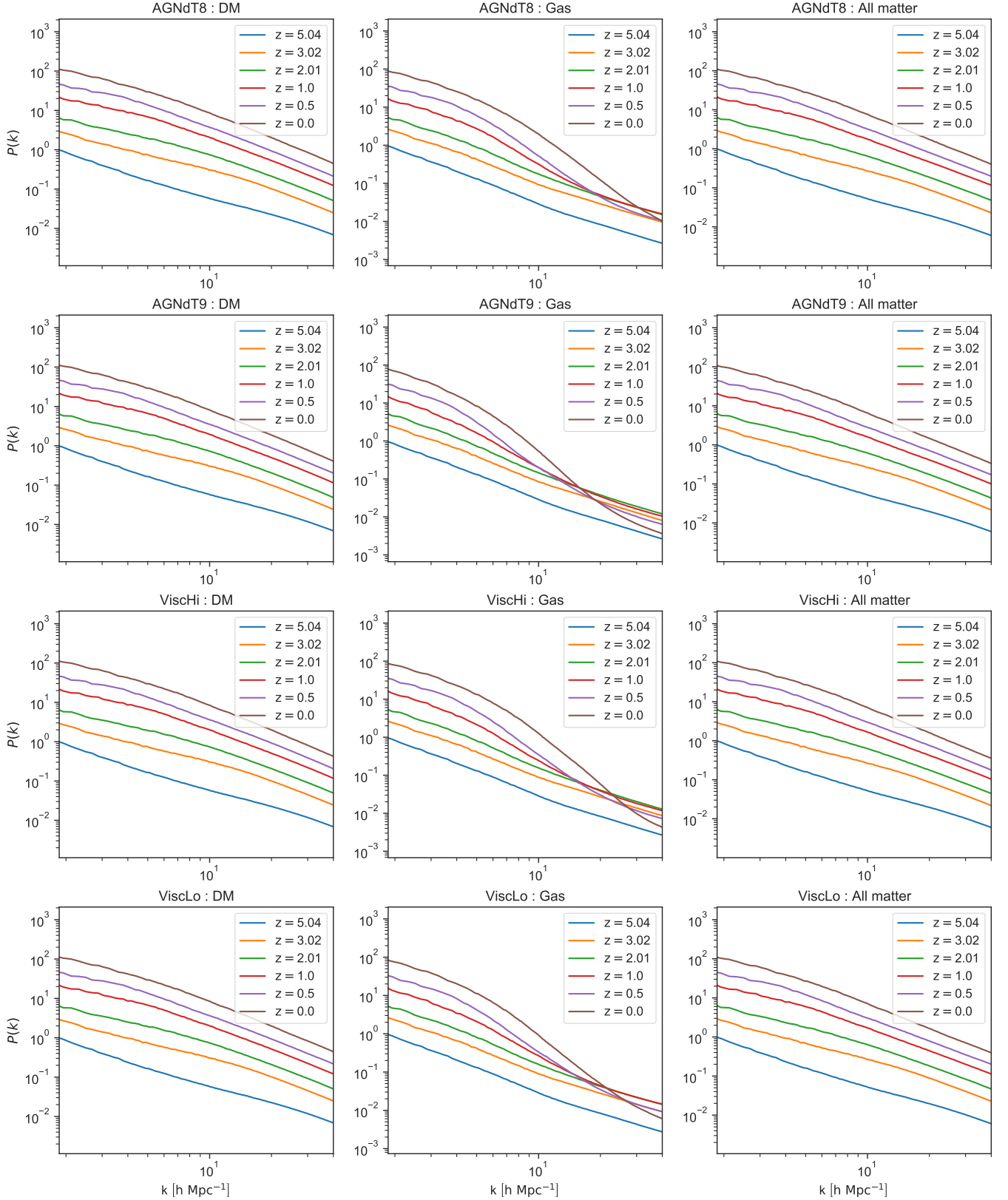


Figure A.2. $P(k)$ for the different models that we studied at different epochs. The crossover events are evident in the gaseous component in all the models. As time goes by, matter starts clustering which boosts the $P(k)$ on all scales; however AGN feedback redistributes the gas, which suppresses the $P(k)$ at small scales, but boosts it at intermediate scales.

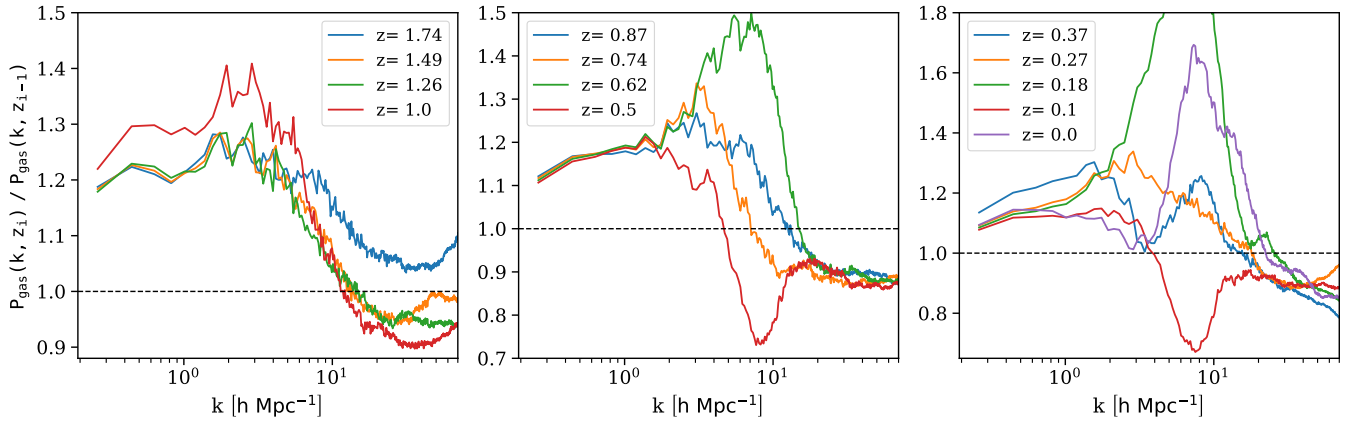


Figure A.3. Ratio of power spectrum of gas between two consecutive epochs for AGNdT9 model. The panels are arranged in decreasing redshift if we go from left to right. The ratio significantly falls below 1 at $z = 1.49$ (first panel from left) which marks the onset of dominance of AGN feedback.

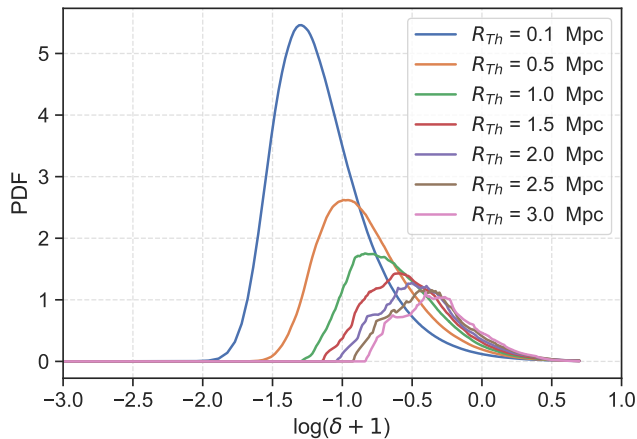


Figure B.1. Distribution of gas at $z=0$ for AGNdT8. The distribution is skewed. It starts to be a mean 0 distribution at very large smoothing scales

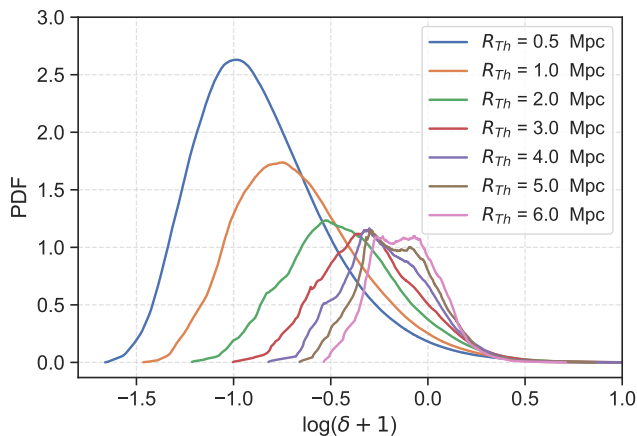


Figure B.2. Distribution of gas at $z=0$ for Eagle Reference Model with Box Size of 100 Mpc. The distribution is skewed. It starts to be a mean 0 distribution at very large smoothing scales, and doesn't differ significantly from the AGNdT8 model.

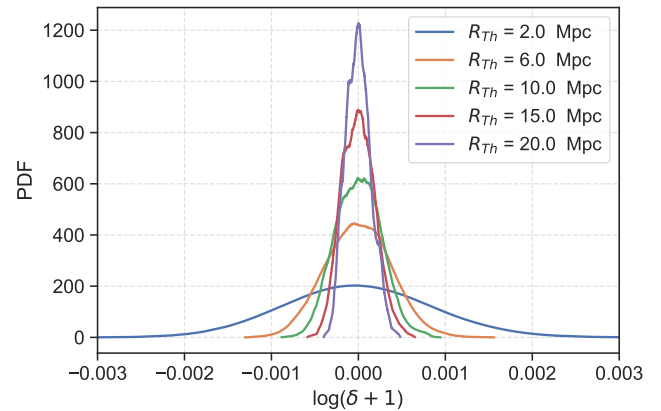


Figure B.3. Distribution of gas at $z=0$ for TNG300-1. The distribution is a very sharp gaussian distribution with mean 0 which is expected for large box size. It becomes a delta function at large smoothing scales.

that we are interested in, such as galaxy cluster and galactic outflows, contains so many particles, that even a 0.1% sample contains enough particle to accurately measure the 3pCF and have a qualitative understanding of the matter distribution. To get the full 3pCF one would need to use the full data set but that poses a significant computational challenge due to a significant increase in the number of particles involved. Unless we have better computational algorithm, it will be very difficult to get the full 3pCF.

We also check the 3pCF matrix for TNG50-2. Qualitatively we observe similar behavior in TNG50-2 also. In Figure-C.4, we compare the 3pCF of AGNdT8 and TNG50-2. The magnitude of 3pCF of TNG50-2 is very comparable to the AGNdT8 model. The differences are attributed to different resolutions and AGN feedback models in the TNG Simulation.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

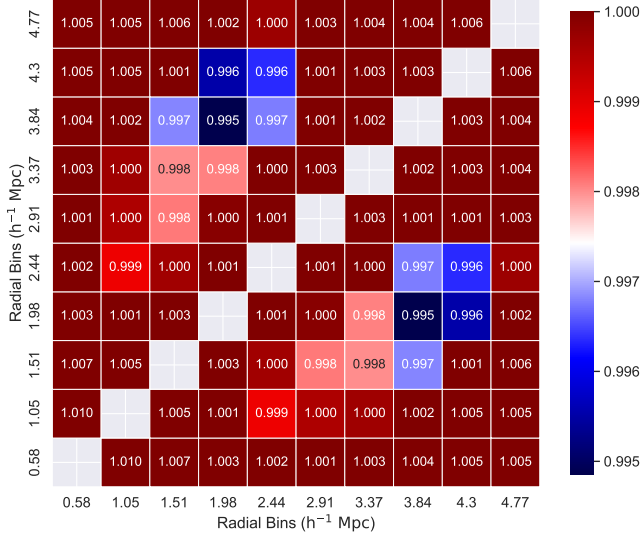


Figure C.1. Ratio of 3pCF between 0.1% and 0.05% subset of the complete data

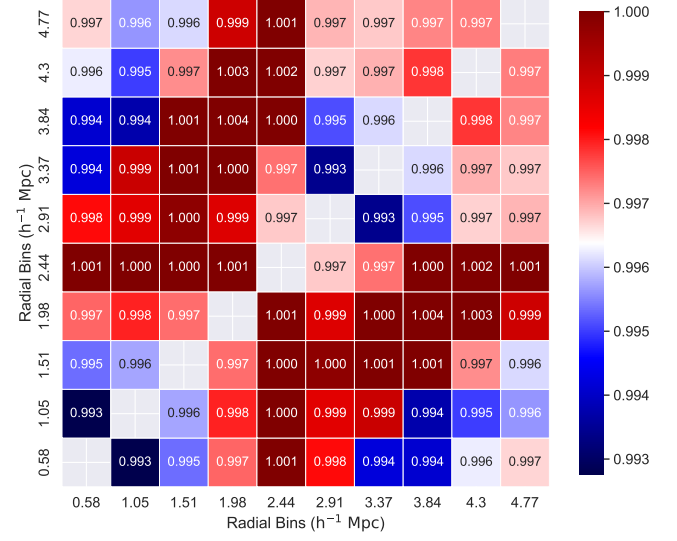


Figure C.3. Ratio of 3pCF between 0.5% and 0.05% subset of the complete data

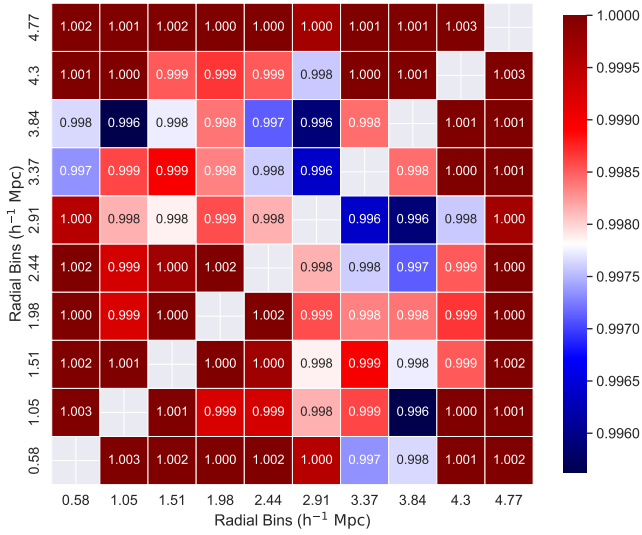


Figure C.2. Ratio of 3pCF between 0.1% and 0.5% subset of the complete data

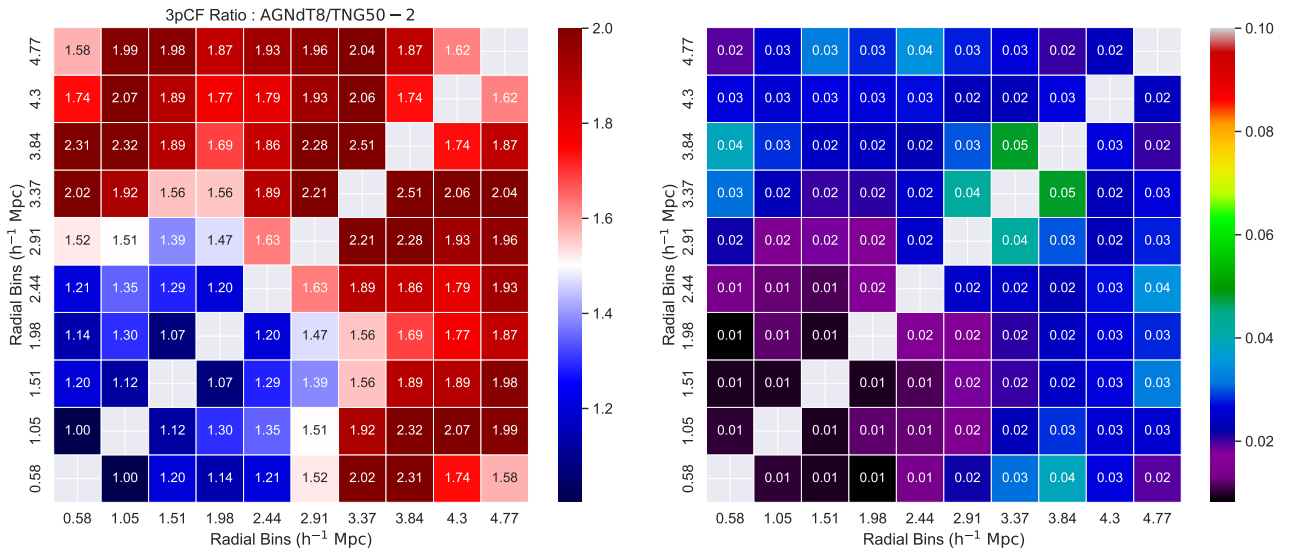


Figure C.4. Comparison of 3pCF between AGNdT8 and TNG50-2. The 3pCF matrix is similar to AGNdT8 and AGNdT9 and qualitatively shows the same behavior. The slight variation is attributed to different resolution of the two simulation and different parameters for AGN feedback.