The formation of supermassive black holes from Population III.1 seeds. II. Evolution to the local universe

N.B. When citing this work, cite the original published paper.

(article starts on next page)

In another scenario, SMBHs are thought to arise from the remnants of the first, so-called Population III (Pop III), stars that formed from the metal free primordial gas at the centre of dark matter 'minihaloes' with masses of $\sim 10^6\,M_\odot$. The remnant black holes might then grow via gas accretion and mergers. There are still significant theoretical uncertainties in the initial mass function (IMF) of Pop III stars (e.g. Haemmerlé et al. 2020). While some numerical simulations predict that a primordial gas cloud will fragment and form multiple low-mass stars (e.g. Greif 2015), other studies have suggested that the gas remains relatively smooth, leading to formation of single or a small multiple of massive stars (e.g. Tan & McKee 2004; Tan, Smith & O'Shea 2010). In the latter case, stellar masses are expected to be $\sim 10^2\,M_\odot$, which would then form black holes of approximately the same mass. However, in any scenario that involves growth from stellar mass black holes there is the fundamental problem that to reach $\sim 10^9\,M_\odot$ scales by $z \sim 7$ requires continuous Eddington-limited accretion, which is challenging to achieve (e.g. Smith, Bromm & Loeb 2017).

A relatively new scenario for SMBH formation has been proposed by Banik, Tan & Monaco (2019), which involves formation of a supermassive star from so-called Pop III.1 sources. These are defined to be the first generation of stars forming from primordial gas at the centre of dark matter minihaloes that are isolated from other astrophysical sources. A potential mechanism to form SMBH seeds with masses $\sim 10^5\,M_\odot$ from such Pop III.1 sources that has been proposed by Banik et al. (2019) invokes the effect of weakly interacting massive particle (WIMP) dark matter annihilation heating, which can delay collapse of the gas cloud until a large mass $\gtrsim 10^5 M_\odot$ builds up. This mechanism requires a high degree of isolation of the minihalo from neighbouring astrophysical sources, which would otherwise provide external sources of feedback, especially ionizing radiation and Lyman-Werner (LW) radiation that heats the gas and destroys H$_2$ and thus disrupts the supermassive star formation process (see also Visbal, Bryan & Haiman 2020). The distance out to which such feedback suppresses Pop III.1 formation is a free parameter of the model, i.e. the isolation distance, $d_{\rm iso}$.

The study of Banik et al. (2019) introduced these ideas and presented a first attempt at their numerical implementation. They predicted SMBH number densities as a function of redshift for a limited range of $d_{\rm iso}$ from 50 to 100 kpc (proper distances). Recently, Singh, Monaco & Tan (2023) (hereafter Paper I) improved the numerical modelling of this scenario and extended the range of $d_{\rm iso}$ that was explored, from 50 kpc to 1 Mpc (proper distances). The properties of the cosmological dark matter density field and halo population were obtained via the PINOCCHIO code (Monaco, Theuns & Taffoni 2002; Munari et al. 2017) that generates dark matter halo catalogues, with known spatial distribution and merger history, based on Lagrangian perturbation theory (LPT). The analysis of Paper I was carried out from redshifts of $z \sim 50$ down to $z = 10$. One of the main results of this study was that SMBHs form in a redshift range from $z \sim 45$ to ~ 20, with a higher value of $d_{\rm iso}$ leading both to formation of fewer SMBHs and to the formation process being completed at a relatively higher redshift, i.e. typically by $z \sim 25$ to 30.

The comoving number density at $z = 10$ was then found to be near constant. Since the values of number density of SMBHs found at $z = 10$ are expected to be similar to those in the local, $z = 0$ universe, the observed local SMBH number density can be used to constrain the Pop III.1 model, especially the isolation distance parameter $d_{\rm iso}$. However, the comparison of the $z = 10$ results with local observations was only approximate, since it did not account for the evolution of the number density from $z = 10$ to 0. One of the main goals of this paper is to address this deficiency by extending the calculation of the evolution of the SMBH population down to $z = 0$.

Extending the model to $z = 0$ will also allow study of several other aspects relevant for comparison with the observed local universe. Not only the overall SMBH number density, but also the occupation fraction of haloes of different masses is a quantity that can potentially be compared with observations. A related property is the fraction of galaxies, especially lower mass dwarf galaxies, that host a central massive black hole. This latter quantity has been studied observationally by, e.g., Miller et al. (2015), Nguyen et al. (2019), and Greene, Strader & Ho (2020), and in the context of several theoretical models by, e.g., Bellovary et al. (2019) and Mezcua et al. (2019).

Another important property of the SMBH population is its clustering, which can be studied via the two-point correlation function. Since in the Pop III.1 scenario the sources are subject to a maximum in their clustering set by $d_{\rm iso}$, this imprints a particular feature in the correlation function. In Paper I, we presented the angular correlation function of the sources and how it evolves with redshift down to $z = 10$. Here, we extend this analysis to $z = 0$ and also consider the full 3D, i.e. spatial, two-point correlation function.

It is important to note that we also make predictions of the evolution of the correlation function considering a different seeding model, i.e. a simple 'halo mass threshold' model, similar to what has been commonly implemented in various cosmological simulations of galaxy formation (e.g. Di Matteo et al. 2008, 2012, 2017; Sijacki et al. 2015; Bhowmick et al. 2022b). These predictions are relevant to test whether it is possible to distinguish the Pop III.1 model from such alternative models of SMBH formation.

The paper is organized as follows. In Section 2, we summarize the numerical methods, including a brief overview of the PINOCCHIO code and the modelling of the two seeding scenarios, i.e. Pop III.1 and halo mass threshold. We present our results in Section 3, including the redshift evolution of SMBH number density and occupation fractions, and the 2D and 3D correlation functions. In Section 4, we present synthetic observational fields, including a case similar to the Hubble Ultra Deep Field. We discuss our results and present our conclusions in Section 5.

Throughout this paper, we adopt a flat ΛCDM cosmology

2 NUMERICAL METHODS

In this section, we first give a brief overview of the main features of the PINOCCHIO code that are relevant for our study. Then we summarize our method of assigning SMBHs to dark matter haloes.

Another mechanism to form intermediate or even SMBHs is through runaway stellar mergers in young and dense clusters to create stars with masses of the order $\sim 200 - 10^3\,M_\odot$ (e.g. Portegies Zwart et al. 2004). This mass can be reached through repeated collisions if the massive stars can reach the cluster core to increase the collision rate drastically (Ebisuzaki 2003) before they explode as supernovae. Gas accretion-driven compression of a dense cluster of stellar mass black holes to form an SMBH has also been proposed (Kroupa et al. 2020). However, in general, predicting whether the conditions needed for such dense clusters arise in galaxies and at what rate is very challenging given the the need to resolve the formation and evolution of individual stars, so predictions for the cosmological population of such systems are highly uncertain (see e.g. Boekholt et al. 2018; Chon & Omukai 2020; Tagawa, Haiman & Kocsis 2020).

Another class of SMBH seeding model considers the very first, so-called Population (Pop) III stars as potential progenitors. However, conventional models of Pop III star formation predict stellar masses that are 'only' $\sim 100\,M_\odot$ (e.g. Abel, Bryan & Norman 2002; Bromm, Coppi & Larson 2002; Tan & McKee 2004; McKee & Tan 2008; Hosokawa et al. 2011; Susa, Hasegawa & Tominaga 2014), which would only have the ability to produce stellar-mass black holes, i.e. relatively low-mass seeds. However, as discussed in more detail below, Pop III SMBH seeding models have been revived by Banik, Tan & Monaco (2019) when allowing for the potential effects of dark matter self-annihilation on the mass scale of formed stars (Spolyar, Freese & Gondolo 2008; Natarajan, Tan & OShea 2009; Freese et al. 2010; Rindler-Daller et al. 2015).

More exotic models involving modification of the standard cold dark matter paradigm have also been proposed. For example, if dark matter undergoes self-interaction, then this could provide a mechanism for SMBH seeding via collapse of the haloes themselves (e.g. Feng, Yu & Zhong 2021). An even more extreme scenario is one in which SMBHs are primordial black holes, although this appears to be disfavoured by the clustering analysis of Shinohara et al. (2023).

Given the uncertainty of SMBH formation models and the difficulty of resolving the small-scale physics, cosmological simulations have typically made very simplified assumptions for the SMBH seeding process based on the properties of the parent halo or galaxy. One of the simplest and most widely used models is the halo mass threshold (HMT) seeding scheme based on the methods developed by Sijacki et al. (2007) and Di Matteo et al. (2008), in which a seed black hole is assumed to form in a halo crossing a certain mass threshold. The Illustris project (Vogelsberger et al. 2014) used this mechanism to add SMBHs of mass $1.4 \times 10^5\,M_\odot$ in each halo, which crosses a mass threshold of $m_{th} = 7.1 \times 10^{10}\,M_\odot$. A similar approach was used in the Evolution and Assembly of GaLaxies and their Environments simulation (Barber et al. 2016).

More recent simulations have taken into consideration additional properties of the host galaxy for SMBH seeding. For example, the Horizon-active galactic nucleus (AGN) simulation (Volonteri et al. 2016) required gas and stellar densities and stellar velocity dispersion to exceed certain thresholds for a galaxy to form a black hole, with a seed mass of $10^5\,M_\odot$ adopted. In addition, all the forming black holes needed to be separated by at least 50 comoving kpc, and their formation was only allowed down to $z = 1.5$. Adopting similar criteria, the OBELISK simulation (Trebitsch et al. 2021) also applied conditions of gas and stellar density needing to exceed certain thresholds, including Jeans instability of the gas, as well as a required isolation of 50 kpc from other SMBHs to avoid multiple black holes forming in the same galaxy. If all these conditions were satisfied, then a black hole of $3 \times 10^4\,M_\odot$ was assigned to the galaxy. In another approach, the ROMULUS simulation (Tremmel et al. 2017)

employed criteria of a limit on metallicity, a threshold on gas density, and a restricted temperature range for SMBH formation, with a seed mass of $10^6\,M_\odot$ adopted. In yet another axample, Bhowmick et al. (2022b) have considered a variety of gas-based SMBH seeding prescriptions and a range of seed masses from $\sim 10^4$ to $10^6\,M_\odot$. While the investigation of certain thresholds of physical quantities for SMBH formation is an advance on a simple HMT models, the above studies are still far from being a complete physical description of SMBH formation.

In this work, we focus on a formation scenario in which Population III.1 stars are the progenitors of SMBHs. Pop III.1 stars are defined to be Pop III (i.e. metal free) stars forming in the first dark matter minihaloes to form in a given region of the universe and so are isolated from other stellar or SMBH feedback sources (McKee & Tan 2008). In this model, it is assumed that in the absence of any significant radiative (or mechanical) feedback, a single dominant protostar forms at the centre of the minihalo and has its structure affected by the energy input from Weakly Interacting Massive Particle (WIMP) dark matter self annihilation inside the protostar (Spolyar et al. 2008; Natarajan et al. 2009; Freese et al. 2010; Rindler-Daller et al. 2015). Such protostars maintain relatively cool photospheres and thus low levels of ionizing feedback, which allows efficient accretion of the baryonic content of the minihalo, i.e. $\sim 10^5\,M_\odot$, to form a supermassive star, which subsequently collapses efficiently to an SMBH after a few Myr.

This Pop III.1 seeding mechanism, which is based on locating isolated minihaloes, was applied in a cosmological simulation by Banik et al. (2019; hereafter Paper I). The evolution was followed from high redshifts down to $z = 10$. The main free parameter in the model is the *isolation distance* (d_{iso}), i.e. how far a newly forming minihalo needs to be from previously formed haloes in order to be a Pop III.1 source. For a fiducial value of $d_{iso} = 100$ kpc (proper distance), the model yields co-moving number densities of SMBHs that match the estimated level of the known $z = 0$ SMBH population. Note that, in this case (and all other reasonable cases), most minihaloes do not form Pop III.1 sources. Rather, most are Pop III.2 sources, which are metal free, but having been disturbed by radiative feedback, they are expected to undergo significant fragmentation to form only lower-mass (e.g. $\sim 10\,M_\odot$) stars (Greif & Bromm 2006).

In this paper, we take this Pop III.1 seeding mechanism and extend the results down to the local universe, $z = 0$. In Section 2, we briefly describe our seeding algorithm and the tools used to apply it. Then we present our results in Section 3, starting with the evolution of number density of seeded haloes down to $z = 0$. We compare these results with the HMT scheme and discuss the SMBH occupation fraction and clustering properties of seeded haloes. Finally, we create synthetic Hubble Ultra Deep Fields (HUDFs) to demonstrate the possibility of using the HUDF to differentiate among different seeding mechanisms. We then present our conclusions in Section 4.

2 METHODS

2.1 PINOCCHIO simulations

As in Paper I, to test our Pop III.1 seeding mechanism, we used the PINOCCHIO code (Monaco, Theuns & Taffoni 2002; Munari et al. 2017) to generate a cosmological box of 59.7 Mpc (40 h^{-1} Mpc for $h = 0.67$) with standard Planck cosmology (Planck Collaboration 2020) and study the formation of DM (mini-)haloes in that box. PINOCCHIO uses Lagrangian Perturbation Theory (e.g. Moutarde et al. 1991) to approximate the evolution of cosmological

perturbations in a ΛCDM universe. For a given set of initial conditions, the code generates outputs in the form of catalogs at different redshifts, which contain mass, position, and velocity of the DM haloes and a complete information of the merger histories of all the haloes with continuous time sampling.

This code was written for applications in cosmology, where huge volumes with moderate mass resolution are requested, and its performance heavily depends on the mass resolution adopted. To resolve minihaloes of $\sim 10^6 \, M_\odot$, it is necessary to sample a 59.7 Mpc box with 4096^3 particles; this results in a particle mass of $1.23 \times 10^5 \, M_\odot$, and we adopted a minimum mass of 10 particles (that would be unacceptable for an *N*-body simulation, but it is acceptable for a semi-analytic code like PINOCCHIO), resulting in a minihalo mass of $1.23 \times 10^6 \, M_\odot$.

Such a large simulation can only be run on a supercomputer, distributing the computation on a large number of MPI tasks. The construction of haloes from collapsed particles is performed in Lagrangian space: the box is divided in sub-boxes, and the grouping algorithm is run on the the particles belonging to its domain. Haloes near or across the sub-box borders would not be constructed correctly, so the sub-box is augmented with a 'boundary layer' (a ghost region) whose size should scale with the Lagrangian radius R_{max} of the largest halo one expects to find in the simulation volume (that can be of the order of several Mpc). This implies an overhead in memory that can be significant. When dividing a small box into many tasks, the size of the sub-boxes can be of the same order of (if not larger than) R_{max}, making the memory overhead unsustainable. The constraint is weakened by stopping the simulation at higher redshift, when R_{max} is still small. As a result, with V4 of PINOCCHIO (Munari et al. 2017) used in Paper I, we were only able to push the simulation down to $z = 10$.

We use here the novel V5 of the code, which implements a number of numerical techniques to improve memory efficiency. This code will be presented elsewhere; the strategy to perform halo construction at high resolution is the following. The sub-box is augmented with a boundary layer as large as needed, but instead of storing the properties of all particles in the augmented sub-box, we start by storing only the particles that lie in the sub-box (excluding the boundary layer) and are predicted to collapse by $z = 0$. Then the halo construction code is run once, collecting a tentative list of haloes; after that, all the particles that are in the boundary layer and lie within N_{Lag} times the Lagrangian size of any formed haloes are added to the list of particles. After collecting the extra information, the halo construction code is run again, generating the final list of haloes. Memory occupation thus depends on the parameter N_{Lag}; our tests show that $N_{Lag} = 3$ guarantees a convergent result, but an extreme run such as the one we present here was possible only by using $N_{Lag} = 2$. The 59.7 Mpc box with full 4096^3 resolution was thus run to $z = 0$ on 800 MPI tasks over 100 computing nodes (each with 256 GB of RAM), so the domain was divided into $6 \times 6 \times 7.5$ Mpc sub-volumes for halo construction. The resulting halo mass function showed two problems that are presented in greater detail in an appendix. We discuss here their nature and their implications.

As a consequence of the difficulty of calibrating the formation of haloes with a very steep power spectrum, the mass of the first haloes is underestimated by a factor of ~ 2 at $z \sim 30$, decreasing to a negligible value at $z \sim 10$. This is a known trend in PINOCCHIO, visible, e.g. in fig. 1 of Munari et al. (2017), where the $z = 3$ halo mass function is slightly underestimated in those tests. We are working to improve this prediction, but we do not consider this as a showstopper for several reasons: our seed BHs are already predicted to form very early, so this underestimation only causes

us to be slightly conservative in their formation redshift, i.e. in fact, they would already have formed at slightly higher z. In our simple modelling, we are assuming here immediate formation of the protostar and then the SMBH, whereas in reality, this might take several Myr or even tens of Myr. The time span that separating $z = 32$ from $z = 29$ is only ~ 14 Myr, so neglecting astrophysical time-scales leads to an overestimation of formation redshift, which compensates against the underestimation problem. Finally, the minihalo threshold mass can be consider to be a second free parameter of the modelling (although one that has physical motivation to be close to $10^6 \, M_\odot$), so one can simply consider our predictions to be valid for minihalo masses of $2.5 \times 10^6 \, M_\odot$. We add to these arguments the fact that inaccuracies in halo masses do not propagate as inaccuracies in halo positions, which are crucial outcomes of our seeding scheme.

A more serious problem is connected to the inaccurate reconstruction of haloes more massive than $10^{12} \, M_\odot$. Indeed, the small size of the sub-box domain for constructing haloes results in a poor reconstruction of massive haloes. This problems makes predictions at $z = 0$ unreliable. We thus produced the same box at a lower resolution, sampled with 1024^3 particles, on a single MPI task on a 256 GB node. Again, this was possible thanks to V5 of the code. In this case, halo construction is as good as it can be. However, the identification of haloes that contain seed SMBHs has been performed in the high resolution box, and though the simulations share the same large-scale structure, matching massive haloes in the two boxes is not a clean procedure. We then resorted to this algorithm: starting from the fact that one low-resolution particle contains 64 high-resolution ones, we calculated which particle in the lower resolution box includes the seeded mini-halo, and assigned the seed to the halo that contains that specific low-resolution particle. We checked that results at $z = 0$ produced with the low- and high-resolution simulations were consistent, with a significant difference in clustering of haloes more massive than a certain threshold that is an expected consequence of the inaccurate mass reconstruction and the known relation of halo bias with halo mass. In the following, we will present results at $z = 0$ based on the low-resolution box, unless mentioned otherwise.

2.2 Seeding scheme

To determine which haloes are seeded with a Pop III.1 star and thence SMBH, consider the scenario depicted in Fig. 1, unfolding in the early universe. The figure shows three stars A, B, and C in different haloes where only A and C become Pop III.1 stars, whereas B is a Pop III.2 star, depending on the separation and formation order. Star A formed first, which then influenced its environment within a sphere of radius equal to $d_{feedback}$, expected to be primarily radiative feedback. Since this star is in a pristine primordial gas without the influence of any feedback from nearby stars, it is defined to be a Pop III.1 star. Star B, which subsequently forms at a distance less than $d_{feedback}$ from star A, is affected by the feedback and hence is a Pop III.2 star (or even a Pop II star if it has been chemically polluted). Finally, star C forms beyond the regions affected by feedback from sources A and B and is thus also assigned to be a Pop III.1 star and thus an SMBH. For the model considered here, the feedback distance is set equal to the isolation distance d_{iso}. So effectively, the condition for a star to be regarded as a Pop III.1 star is that when it is forming, there should be no previously formed haloes present in the sphere of radius d_{iso}. We consider d_{iso} as a free parameter in our theory and vary it to match the observed number density of the SMBHs in the local Universe.
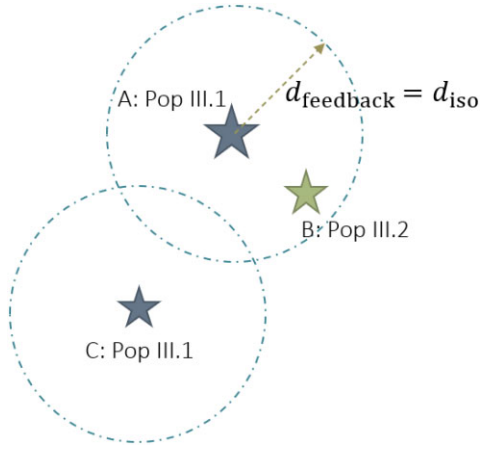
Figure 1. A schematic illustration of the Pop III.1 SMBH seeding scenario depicting the conditions for a star to be isolated enough to be considered as a Pop III.1 star (see text).

2.3 Seed identification in the dark matter catalogs

To perform the seed identification analysis from the dark matter catalogs generated by PINOCCHIO, we first divided the entire redshift range (from $z = 0$ to the redshift when the first minihalo forms, $z \approx 40$) into small bins of widths ranging from $\Delta z = 1, 2,$ or 3, depending on the output catalogs available, which in turn depends on the relative change in positions of (mini)haloes. The bins are wider at high redshifts, but smaller at lower redshifts. Then for each redshift interval (z_l, z_h) where $(z_h > z_l)$, we utilized k-d tree data structure to create a three-dimensional map in position space of all the haloes existing between z_h and z_l. The positions used to create the tree are taken from the output catalog of PINOCCHIO at the lower redshift of the interval (z_l). Since the positions are not updated once the tree is constructed, we account for the change in the positions within this redshift interval by finding the maximum change (δ) of position among all the haloes existing for the entire redshift range. Then, for each minihalo crossing the mass threshold of $10^6\,\mathrm{M_\odot}$ (or as in the nomenclature of PINOCCHIO: 'appearing') at a redshift z_{app} $\in (z_l, z_h]$, we perform a ball search using the k-d tree to find all the haloes around the appearing minihalo within a sphere of radius d_{iso} $- 2\delta$.[1] If there exists even a single halo at the redshift z_{app} within this sphere, then this minihalo is flagged as a halo containing a non-Pop III.1 star at its centre. If there are no haloes existing at this redshift, then the ball search is performed again with the same minihalo at the centre, but this time within a sphere of radius $d_{iso} + 2\delta$. Then for all the haloes existing at redshift z_{app} within the shell of radius $d_{iso} \pm 2\delta$, we find the exact distance between the minihalo at the centre and all these haloes using the exact positions at z_{app}. If this distance is greater than d_{iso} for all the haloes within the shell, then the minihalo at the centre is flagged as a Pop III.1 source, i.e. an SMBH-seeded halo. This process is repeated for each minihalo crossing the threshold mass within the two redshifts, and then this whole procedure is performed again for all the redshift intervals until the whole redshift range is covered. In this way, we are able to check the isolation condition for each minihalo appearing in the cosmological box and find all the seeded minihaloes.

---

[1] A factor of 2 is multiplied with δ to account for the change in position of both the minihalo at the centre of the sphere and all the other haloes within the sphere.

At smaller redshifts, the change in positions of the haloes (δ) within the redshift intervals becomes comparable to the isolation distance. This implies that the quantity $d_{iso} - 2\delta$ can become negative (in our simulation box, this happens at around $z \approx 15$ for $d_{iso} = 50$ kpc). In this case, the ball search is directly performed in a sphere of radius $d_{iso} + 2\delta$, and then the exact distances between the minihalo at the centre and all the other haloes existing at z_{app} are calculated.

This division of the entire redshift interval and creating the k-d only at specific redshifts is performed to avoid reconstructing the tree with the up-to-date position at every instance a new minihalo appears. Since the number of minihaloes is very large, it becomes highly expensive computationally to reconstruct the tree with updated positions each time a new minihalo appears.

3 RESULTS

3.1 Number density evolution

As explained in the last section and in Paper I, we identify SMBH-seeded haloes by the condition that the isolation sphere of radius d_{iso} around a newly forming minihalo is not populated by any other existing halo (of mass greater than our minihalo threshold mass). The obtained results for the evolution of number density for different values of d_{iso} (in proper distance units) are shown in Fig. 2. The coloured dotted lines show the number density evolution of total number of SMBHs, whereas the coloured solid lines show the number density for seeded haloes (which are slightly smaller, especially at lower redshifts, due to mergers). Compared to the number densities in fig. 1 of Paper I, the values obtained here are moderately lower (by a factor of 1.45 for 100 kpc and 1.65 for 50 kpc) because we have considered periodic boundary conditions when identifying the seeds, which was not done in Paper I.

Fig. 2 also shows some observational estimates of n_{SMBH}. An estimate at $z = 0$, presented in Paper I, is calculated by assuming that each galaxy with luminosity greater than $L_{min} = 0.33L_*$ hosts a SMBH, with the error bar around this point assuming a range of L_{min} from 0.1 to 1.0 L_*. Note, L_* is the characteristic luminosity corresponding to $M_B = -19.7 + 5\log h = -20.55$ (e.g. Norberg et al. 2002). Recent observations of high redshift AGNs from JWST surveys have started providing lower limits on the number density of SMBHs in the early universe, with one such estimate presented by Harikane et al. (2023), from the sample of Nakajima et al. (2023; black diamonds in Fig. 2). Their estimate provides lower bounds on the observed number density of Type I AGN at redshifts $z = 4$–7.

Fig. 2 illustrates the expected behaviour that as the isolation distance is reduced, the number of formed SMBHs increases, i.e. it is easier to satisfy the isolation distance criterion. We can also conclude that for a certain range of d_{iso} (≈ 90–170 kpc), the number density obtained is in reasonable agreement with the $z = 0$ estimate. Thus, the case with $d_{iso} = 200$ kpc is disfavoured simply by its inability to produce enough SMBHs. A key feature of the fiducial model, i.e. with $d_{iso} = 100$ kpc, is that *all SMBHs have formed very early in the Universe: the process is essentially complete by* $z \simeq 25$.

Fig. 2 also shows results for an example HMT model (shown by green dashed line) in which each halo more massive than $m_{th} = 7.1 \times 10^{10}\,\mathrm{M_\odot}$ is seeded (e.g. the Illustris project: Vogelsberger et al. 2014; Sijacki et al. 2015, etc.); note, this seeding scheme is driven by the mass resolution of the simulation, i.e. haloes are seeded as soon as they are resolved with a sufficient number of particles). The main difference compared to the fiducial Pop III.1 model is in the overall number density of SMBHs at $z \gtrsim 5$.
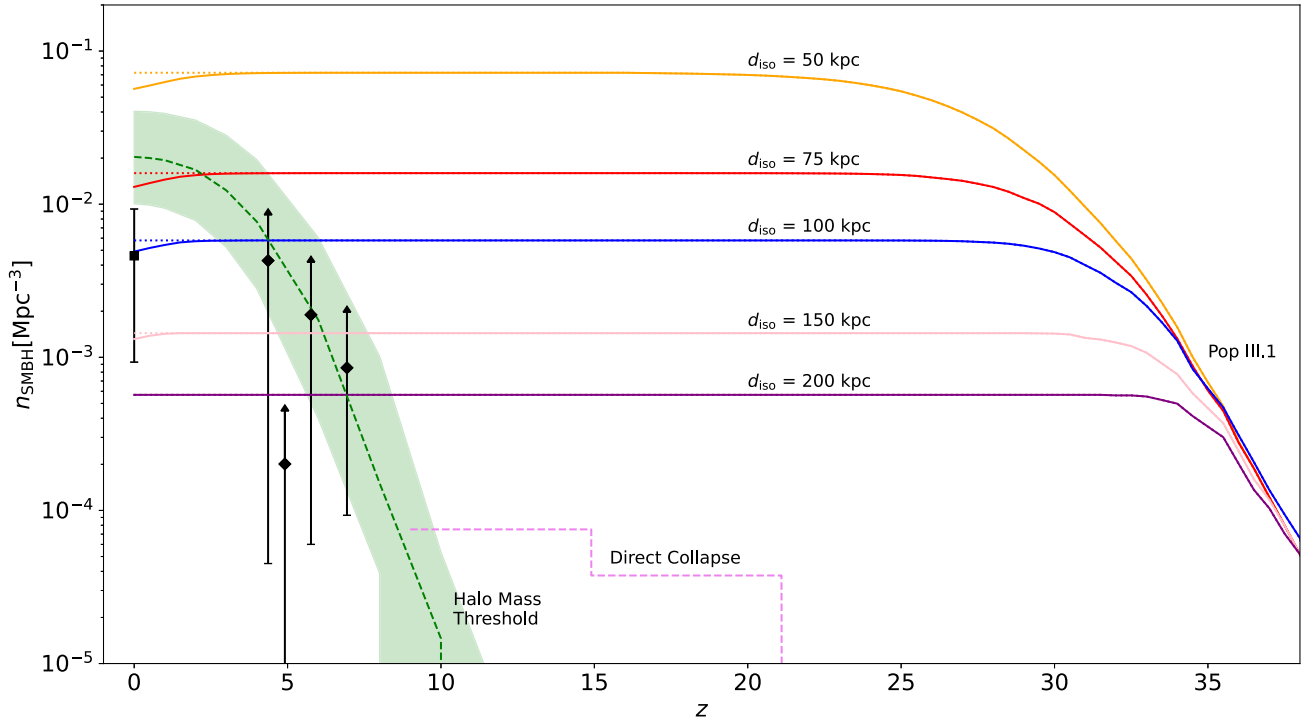
Figure 2. Evolution of the comoving number density of SMBHs, n_{SMBH}, for different theoretical models. Results for Pop III.1 models with several values of isolation distance (in proper distance) are shown, as labelled. The dotted lines show the total number of SMBHs that ever formed, while the solid lines show remaining number of seeded haloes after accounting for mergers. An example HMT model is shown by the dashed green line in which each halo with mass higher than $m_{th} = 7.1 \times 10^{10}$ M$_\odot$ is seeded (see text). The green shaded region shows the effect of lowering and raising m_{th} by a factor of 2. The violet dashed line shows the results of a simulation modelling SMBH formation via direct collapse (Chon et al. 2016). The black solid square indicates an estimate for the number density of SMBHs at $z = 0$, assuming each galaxy with luminosity higher than $L_{min} = 0.33L_*$ contains one SMBH, with the range shown by the error bar obtained by varying L_{min} from $0.1L_*$ to L_*. The black diamonds are estimated lower limits of n_{SMBH} from JWST observations of Type I AGN (Harikane et al. 2023).

Table 1. Total number of formed SMBHs ($N_{SMBH, form}$), total number of SMBHs remaining at $z = 0$ assuming efficient mergers [$N_{SMBH}(z = 0)$], the difference between these [$\Delta N_{SMBH} = N_{SMBH, form} - N_{SMBH}(z = 0)$], which is equivalent to the number of mergers, and the fraction of original SMBHs that are destroyed by mergers ($f_{merger} = \Delta N_{SMBH}/N_{SMBH, form}$).

d_{iso} (kpc)	$N_{SMBH, form}$	$N_{SMBH}(z = 0)$	ΔN_{SMBH}	f_{merger}(%)
50	15 356	12 051	3305	21.52
75	3394	2760	634	18.68
100	1234	1043	191	15.48
150	306	280	26	8.50
200	121	116	5	4.13

We also show the results of a simulation by Chon et al. (2016) modelling the formation of SMBHs via the direct collapse mechanism. Here, they simulated a $20h^{-1}$Mpc box and found only two SMBHs formed (at $z \simeq 15$ and 21). Even though this simulation was only run down to $z = 9$, the number density is not expected to increase much at lower redshifts, given the conditions assumed to be needed for direct collapse, i.e. massive, irradiated, tidally stable, metal-free haloes. While this model allows some SMBHs to form relatively early, as discussed in Section 1, the overall number densities achieved by this mechanism are much smaller than are needed to explain the entire observed SMBH population.

We quantify the number of mergers that occur between seeded haloes in the Pop III.1 models. Table 1 shows the total number of SMBHs that formed ($N_{SMBH, form}$) and the number of haloes

containing them at $z = 0$ [$N_{SMBH}(z = 0)$]. Assuming efficient merging of SMBHs that are in the same halo, then the number of mergers is $\Delta N_{SMBH} = N_{SMBH, form} - N_{SMBH}(z = 0)$. A feature of the Pop III.1 seeding mechanism is that SMBHs are initially spread out from each other, so that there are relatively few binary SMBHs and few mergers. A detailed analysis of the mergers, including the binary (and higher order multiples) AGN number densities and the gravitational wave background emanating from these mergers, will be discussed in a future paper in this series.

A caveat of our seeding model is that at small redshifts, around $\lesssim 6$, the isolation distance in comoving units becomes so small that many minihaloes that appear after this redshift start satisfying the isolation criteria. This effect would result in an increase in number density by around two orders of magnitude by $z = 0$ from the converged values around $z \approx 20$, for all cases of d_{iso}. However, since reionization has completed by $z \approx 8$ (Planck Collaboration 2020), we assume that the formation of Pop III.1 sources is also not possible below this redshift. Hence, in our analysis, we set a limit of seed formation to be only possible until $z = 8$. For most cases of the isolation distances, we considered (≥ 75 kpc), the number density is already converged at redshifts greater than $z = 20$. However, for the case of 50 kpc, new seeds still keep on appearing until $z = 8$ (although below $z = 15$ the total number only increases by about 1 per cent).

In Fig. 3, we show a visual representation of the seeded haloes in the box at different redshifts, for all the isolation distances considered in Fig. 2. As discussed, the 50 kpc case is the most crowded with the highest number of seeded haloes at every epoch shown. Initially,
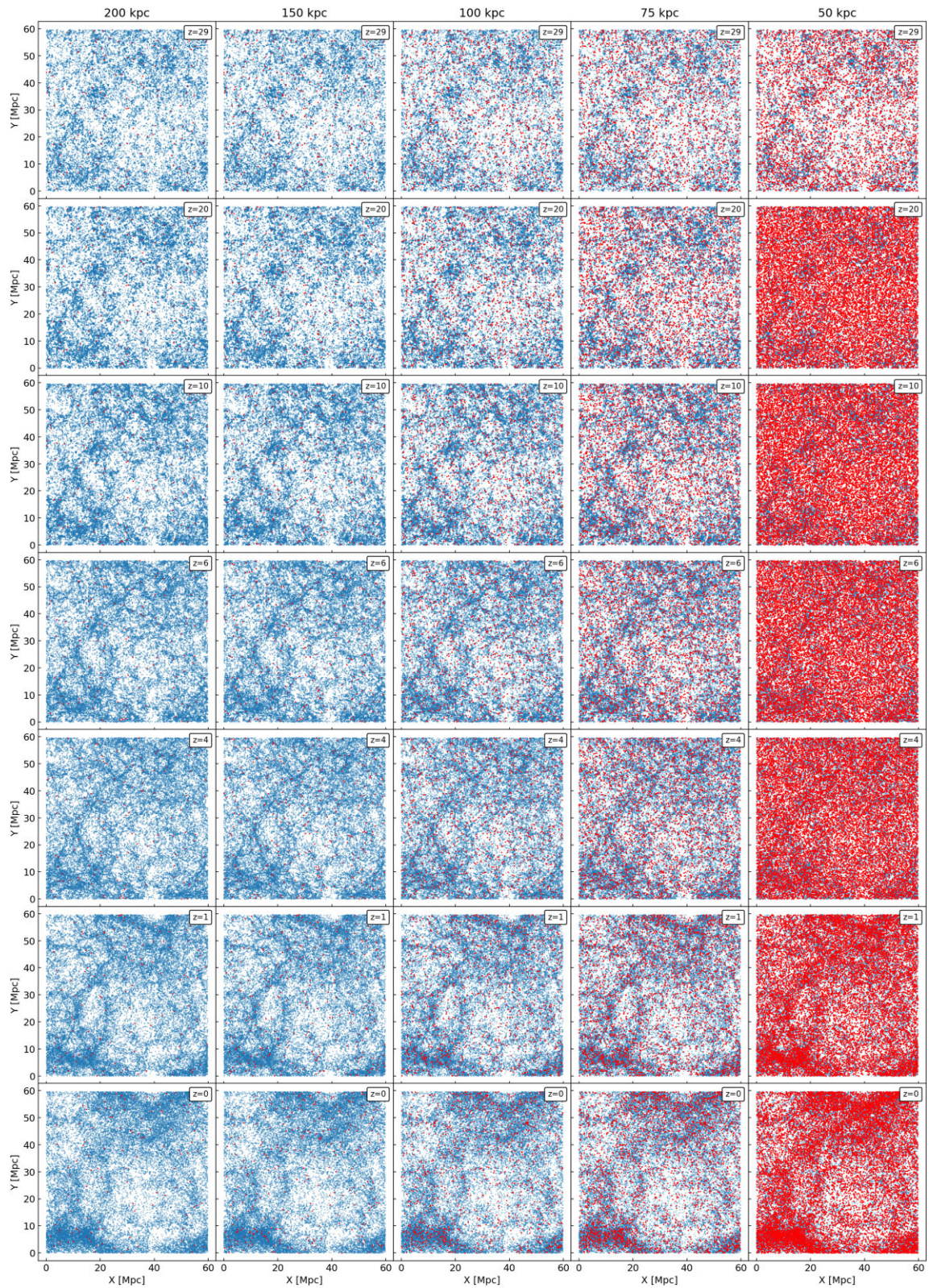
Figure 3. Projection of the positions of seeded haloes (*red*) and non-seeded haloes (*blue*) in the XY plane of the box for different isolation distances. The redshift is shown in the top right corner of each panel (same for each row). Only the 30 000 most massive non-seeded haloes within each panel are shown for ease of visualization.
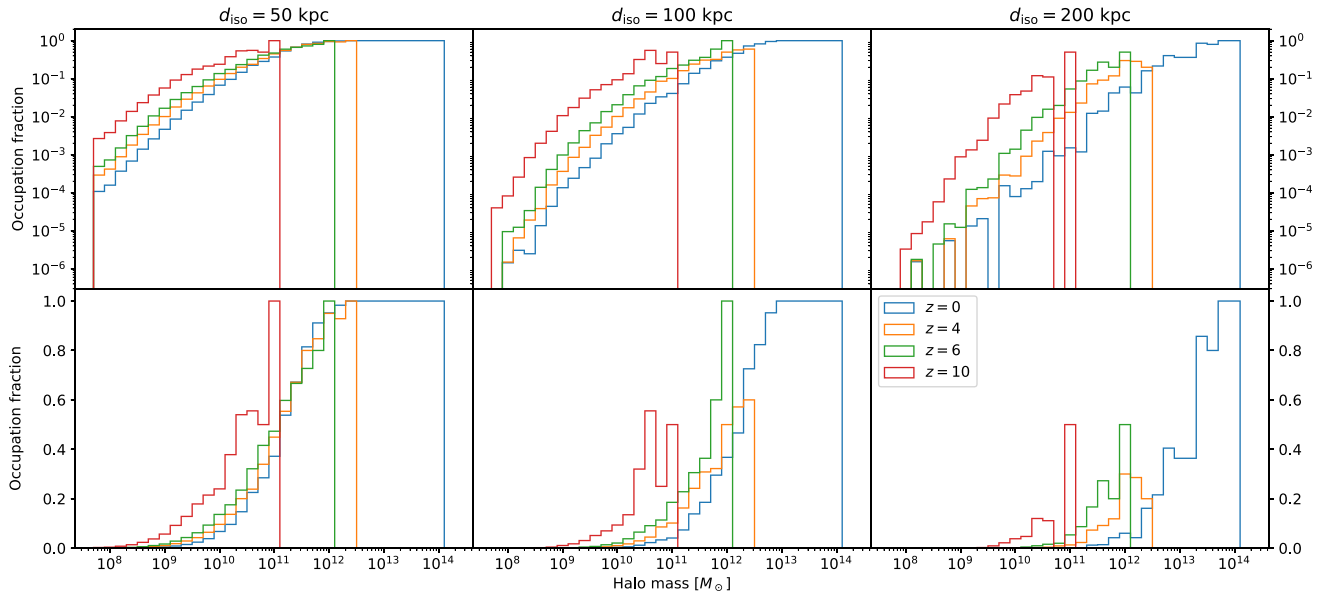
Figure 4. Evolution of SMBH occupation fraction of haloes for different cases of $d_{\rm iso}$. Top row depicts the fraction in log scale, while the bottom row shows the same data in linear scale. The mass bins are divided into equal bins of width 0.2 dex.
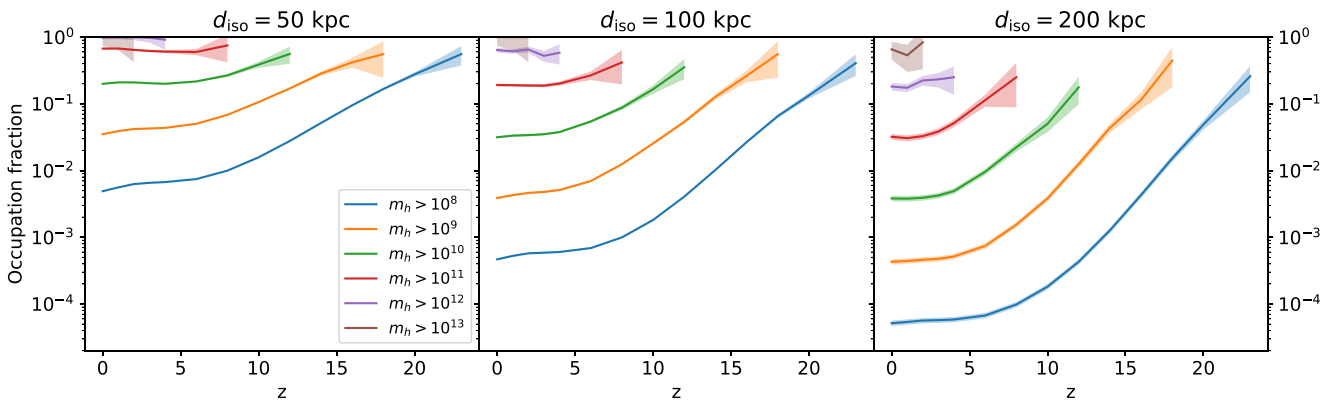
Figure 5. Cumulative occupation fractions of haloes having masses greater than a given value (see legend). The shaded region represents $\pm 1\sigma$ error due to counting statistics.

all the seeds emerge in a relatively unclustered manner, but eventually, the clustering increases as lower-mass seeded haloes migrate towards more massive haloes and merge with them in overdense regions. We perform a more detailed analysis of clustering in Section 3.3.

3.2 Occupation fraction of seeded haloes

From observations of local galaxies, it appears that almost all massive galaxies contain a nuclear SMBH. This implies that the SMBH occupation fraction of haloes should approach unity as halo mass rises. Fig. 4 shows the evolution of occupation fraction from one realization of our 59.7 Mpc box, through four different redshifts for haloes ranging from $[5 \times 10^7, 2 \times 10^{14}] \, {\rm M}_\odot$ (the upper limit of the mass range is chosen to include the most massive halo at $z = 0$ in our 1024^3 resolution simulation box, measuring $1.2 \times 10^{14} \, {\rm M}_\odot$). As expected, with the decrease in the isolation distance, more and more haloes are seeded and hence the occupation fraction is higher compared to the same mass range for larger $d_{\rm iso}$. All the fractions

at $z = 0$ approach unity for the most massive haloes, independent of the isolation distance. Interestingly, the most massive halo is not always occupied by an SMBH throughout the redshift evolution in our simulations. For example, at $z = 4$, there can be significant fractions of the most massive haloes, i.e. $\sim 10^{12} \, {\rm M}_\odot$, that are not seeded, as in the case of $d_{\rm iso} = 100$ kpc. Fig. 4 also shows that for $d_{\rm iso} = 200$ kpc the occupation fraction for haloes with masses $\sim 10^{12} \, {\rm M}_\odot$ at $z = 0$ is quite small, $\lesssim 0.1$, which is a further indication that it produces too few SMBHs.

Fig. 5 shows the evolution of the cumulative occupation fraction, i.e. for all haloes more massive than $\{10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{13}\} \, {\rm M}_\odot$, for three different cases of isolation distance. If we consider only the most massive haloes ($> 10^{13} \, {\rm M}_\odot$), the fraction is close to one (as also evident from Fig. 4). At a given redshift, as we consider less massive haloes, the occupation fraction decreases. At a given mass threshold, as we move out to higher redshift, the occupation generally rises, since these haloes become relatively more extreme members of the global halo population. Interestingly, the occupation fraction for all haloes
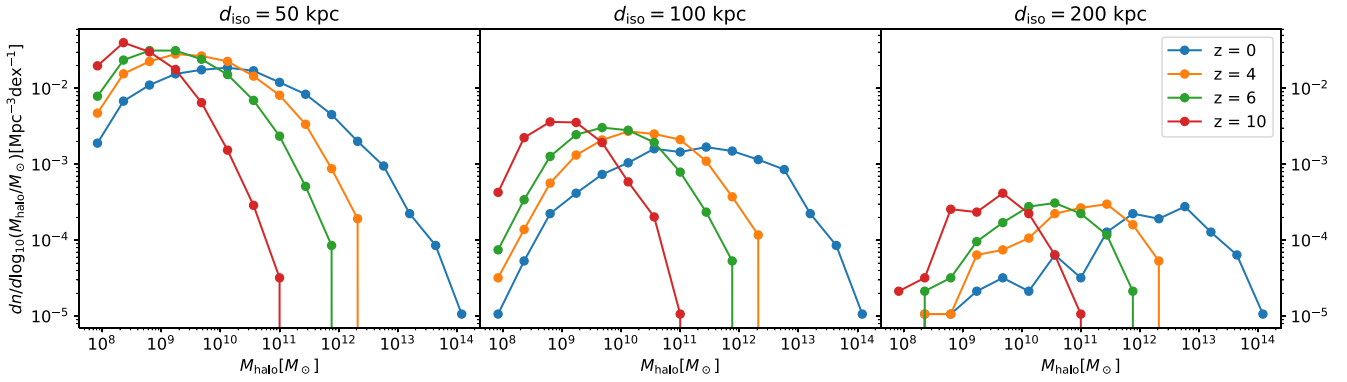
Figure 6. Mass function of seeded haloes at different redshifts for $d_{\rm iso} = 50$, 100, and 200 kpc cases (left to right).
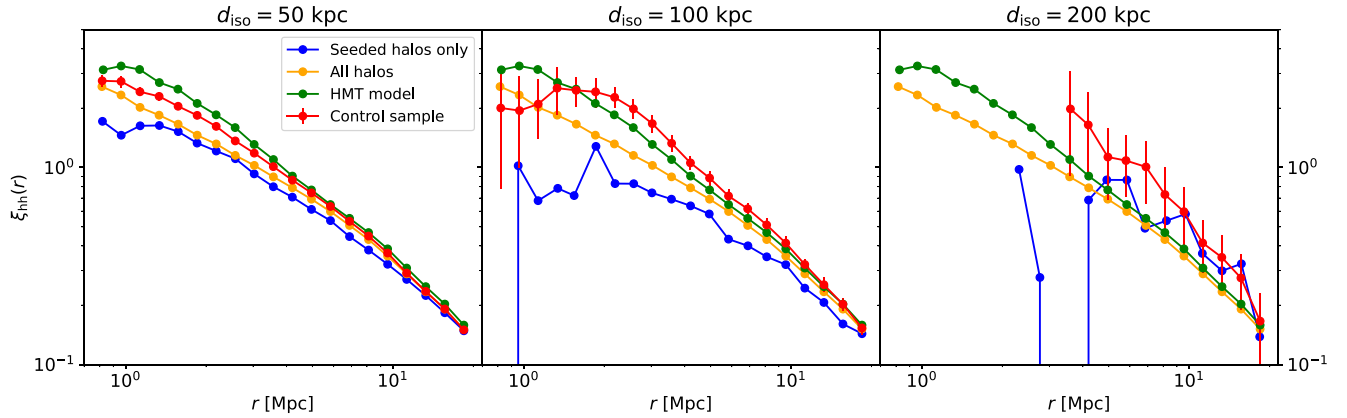

Figure 7. The 3D 2pcf for the seeded haloes more massive than $10^{10}\,\rm M_\odot$, at $z = 0$ for different isolation distances. The blue points show the correlation function for only the haloes containing SMBHs, while the orange points show the correlation for all the haloes, with or without an SMBH. For the red points, we randomly select haloes from the pool of all the haloes, but with the same number and mass distribution as the seeded haloes. The error bars indicate 1σ deviations from the mean value from randomly sampling 50 times. The green points show the correlation for haloes seeded according to the HMT scheme, in which all the haloes greater than $m_{\rm th} = 7.1 \times 10^{10}\,\rm M_\odot$ are seeded.

more massive than 10^8 and $10^9\,\rm M_\odot$ ($10^{10}\,\rm M_\odot$ as well, although to a lower degree) at $z = 0$ differ by factors of approximately 10 among the three cases of isolation distances considered, reflecting the same differences in the global number densities at $z = 0$ (see Fig. 2).

To obtain a better understanding of the mass function of the seeded haloes, in Fig. 6, we present the distribution functions of these haloes for the $d_{\rm iso} = 50$, 100, and 200 kpc cases, including their evolution with redshift. We see that, as expected, these mass functions evolve to higher masses as the universe evolves from $z = 10$ down to $z = 0$. The peak of the seeded halo mass function is lower for smaller values of $d_{\rm iso}$. However, the distributions are quite broad, indicating significant fractions of SMBHs in relatively low-mass haloes, even at $z = 0$. In a future paper in this series, these seeded halo mass functions and the properties of their host galaxies will be compared to SMBH census data, especially focusing on properties derived in the local universe.

3.3 Clustering

We perform a clustering analysis using the CORRFUNC library (Sinha & Garrison 2020) for PYTHON, and the results are shown in Fig. 7. By sampling r in 20 logarithmic bins of $r_{\rm min} = 0.5$ Mpc $\rm h^{-1}$ to $r_{\rm max} = 13.3$ Mpc $\rm h^{-1}$, we evaluate the 3D, two-point correlation

function[2] (2pcf) $\xi_{\rm hh}(r)$ for all haloes more massive than $10^{10}\,\rm M_\odot$ at $z = 0$. Since PINOCCHIO only evolves dark matter haloes, the information of substructures such as subhaloes within haloes is not stored or tracked. This implies that only radial scales greater than the size of a typical dark matter halo (3 to 4 Mpc at $z = 0$) are relevant for consideration. In other words, the correlation function presented here does not include the one-halo term. From the figure, we observe that the clustering of the SMBH-seeded haloes (blue points) is always lower compared to other cases. This is expected because of the nature of our model, which results in larger distances between SMBHs and hence smaller clustering amplitude. The plots for $d_{\rm iso} = 50$ and 100 kpc clearly depict this, while the case of 200 kpc suffers from low number statistics. The red points, which represent the clustering of random haloes with the same number and mass distribution as of the seeded haloes, are generally more than 1σ higher than the blue points, except at the largest scales. This can be clearly seen for the fiducial case of 100 kpc. We also show the clustering for the fiducial case of HMT schemes with $m_{\rm th} = 7.1 \times 10^{10}\,\rm M_\odot$ (Sijacki et al. 2015) depicted by green points. This model also generally

---

[2]All the correlation functions presented in this section have been corrected by analytically adding large-scale clustering modes corresponding to scales larger than the box size. Refer to Appendix B for more details.
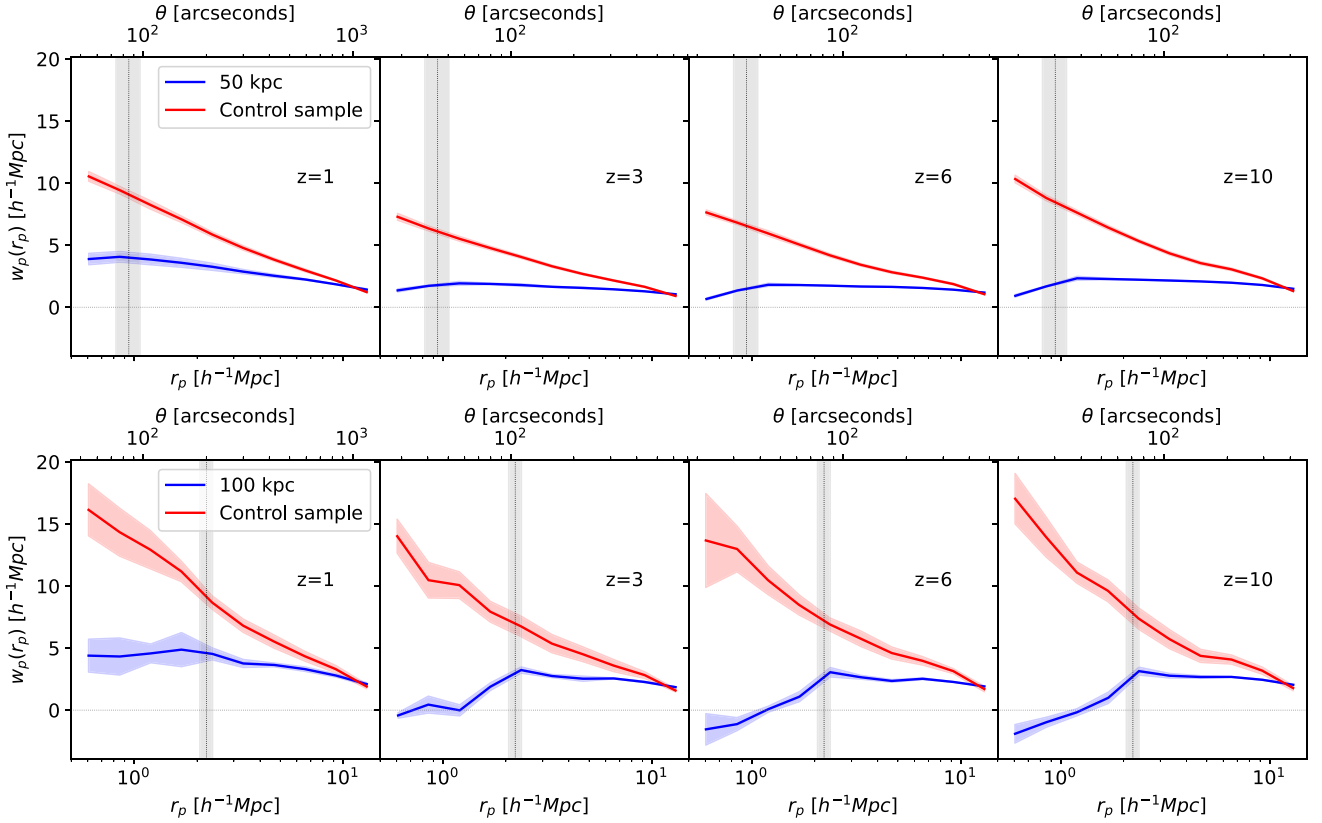
Figure 8. Evolution of projected correlation function for $d_{iso} = 50$ kpc (top row) and 100 kpc (bottom row) cases. The blue line is the average after computing the correlation of the seeds from three orthogonal sides of the box and the shaded region represents the 1σ spread. The control sample is the correlation of haloes selected randomly but with the same mass and number distribution as the seeded haloes at that redshift. The red line refers to the average after randomly sampling 10 times and the shaded region refers to 1σ deviations from the mean. The vertical grey line refers to the size of the isolation radius at the mean formation redshift ($d_{iso}(\bar{z}_{form})$) of the seeded haloes, and the grey region represents 1σ deviation from the mean. For 100 kpc, $\bar{z}_{form} = 32.08$, and for 50 kpc, $\bar{z}_{form} = 27.14$. The angular axis on top of each panel corresponds to the angular scale of r_p projected on the sky at the respective redshift.

shows higher clustering than our Pop III.1 seeding model. Thus a clustering analysis of census of a local Universe ($z = 0$) survey of all (or a significant fraction) of SMBHs has the potential to distinguish between these SMBH seeding mechanisms.

In Fig. 8, we show the evolution of the projected correlation function for the $d_{iso} = 50$ and 100 kpc cases (blue lines), compared to haloes with the same mass and number distribution as the respective seeded haloes (red lines). As seen in the 3D 2pcf, the clustering of the seeded haloes is always lower than the randomly selected haloes and this trend is observed even at higher redshifts. Furthermore, there is a significant drop of the clustering amplitude of the seeded haloes for scales lower than $d_{iso}(\bar{z}_{form})$ (vertical grey band), a signature of feedback cleared bubbles, first discussed in Paper I for $z \geq 10$. Here we see that this signature of suppressed clustering persists to lower redshifts, although is gradually diminished as the Universe evolves to a more clustered state.

We emphasize that comparing our clustering predictions at redshifts greater than 1 or 2 is not feasible with currently available observational data. The measurements from a range of luminosity of AGNs at these redshifts imply minimum halo masses of $\sim 5 \times 10^{11} h^{-1} M_\odot$ at $z \sim 3$ (Allevato et al. 2014) to more than $10^{12} h^{-1} M_\odot$ at $z \sim 4$ (He et al. 2018). For our 59.7 Mpc box, the number of seeded haloes above these thresholds are quite low. For instance, for the $d_{iso} = 100$ kpc case, only around 6 per cent of sources are above this threshold at $z = 3$ and only 0.7 per cent sources are

more massive than $10^{12} h^{-1} M_\odot$ at $z = 4$. If we apply these halo mass cuts on our seeded haloes, then the clustering signal is too noisy to make any decent comparison with the observational data. Moreover, at high halo masses, the occupation fraction approaches unity, so for the measured clustering of bright AGNs, hosted in relatively massive haloes, we expect that they may cluster as their host haloes, with no appreciable difference with respect to currently used models. More data on AGN, especially those that are present in lower-mass haloes/galaxies, are needed to test the models.

As a crude comparison, in Fig. 9, we include the clustering measurements from Zehavi et al. (2011), who performed the projected clustering analysis of volume-limited sample of 570 000 galaxies from the Seventh Data Release (Abazajian et al. 2009) of the Sloan Digital Sky Survey (York et al. 2000). The galaxies used in their data extend out to $z = 0.25$, with a median redshift of $z \sim 0.1$. We compare our results at $z = 0$ for $d_{iso} = 50$ and 100 kpc, along with the HMT scheme, with their galaxy luminosity threshold cut result for $M_r < -19.0$. We computed the relation between DM halo mass and r-band absolute magnitude by comparing the clustering amplitude of PINOC-CHIO DM haloes with Zehavi et al.'s measurements, minimizing the χ^2 of the clustering amplitude only for $r_p > 3h^{-1}$ Mpc (to avoid the one-halo clustering scales); for $M_r < -19.0$, we find a clustering-matched halo mass of $M_{PIN}^{-19.0} = 1.91 \times 10^{12} h^{-1} M_\odot$, higher than the value suggested in that paper ($M_{zehavi}^{-19.0} = 2.55 \times 10^{11} h^{-1} M_\odot$); this

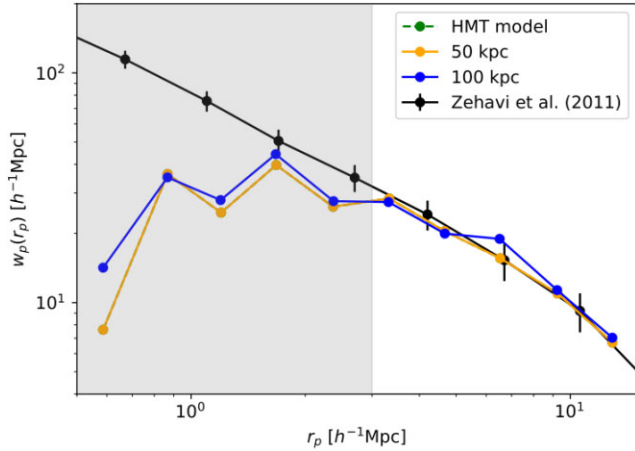**Figure 9.** Comparison of the results for the projected correlation function $w_p(r_p)$ obtained from our simulations for d_{iso} =50 and 100 kpc and the HMT scheme at $z = 0$ with the observational data from Zehavi et al. (2011) for a $M_r < -19.0$ magnitude cut. The shaded region shows scales smaller than the size of a typical halo at $z = 0$, i.e. $r_p < 3h^{-1}$Mpc, which are not of interest for our comparison due to limitations of our model (lack of sub-haloes). The HMT scheme and 50 kpc models overlap, as all haloes above the threshold are seeded for that value of d_{iso}.

is not surprising, given the different cosmology assumed in 2011. We then applied this halo mass cut on our $d_{iso} = 50$ and 100 kpc sources, as well as the HMT scheme, and compared the projected correlation function for the $M_r < -19.0$ threshold galaxies in Fig. 9. For the region of interest, the clustering of the seeded haloes shows good agreement, within the errors, with the observations. The $d_{iso} = 50$ kpc correlation completely overlaps the HMT one because all the sources more massive than $M_{PIN}^{-19.0}$ are seeded in this model. Also, at this high-mass cut, most of the $d_{iso} = 50$ kpc sources are also seeded in the $d_{iso} = 100$ kpc model, and hence their clustering follows similar trends. This is due to the fact that the occupation fraction approaches unity for the most massive haloes (see Section 3.2) for all the isolation distances, and since the mass cut is high, this means

that most, if not all, the haloes are seeded, regardless of the isolation distance.

3.4 Ultra deep field

One potential way to compare our model with observational data is to count the number of SMBHs (i.e. appearing as AGN) present in projected deep fields of the Universe, such as the HUDF (Beckwith et al. 2006; Ellis et al. 2013). We thus create a synthetic ultra deep field (UDF) populated with SMBHs that have formed in our simulations. To achieve this, we use snapshots of haloes at different redshifts in the 59.7 Mpc cosmological box, using the highest resolution run. We pierce the box orthogonally from random positions (avoiding repetitions) and then stack the fields in redshift space to generate the light cone of a 2.4 arcmin side length (i.e. same as the HUDF). Fig. 10 shows our constructed HUDF, for $d_{iso} = 50$ and 100 kpc. The fields shown are for the redshift range $z \in [4, 16]$, with the number of haloes in the field equal to 9352 and 764 for $d_{iso} = 50$ and 100 kpc, respectively. As expected, the field for the 50 kpc case is much more densely populated with seeded haloes as compared to 100 kpc.

Fig. 11 shows the distribution of SMBHs within the redshift range $z = 5$–10 in our synthetic HUDF, where we also display the number of sources in redshift bins of $\Delta z = 1$. The total number of sources in the field (*last column*) for the fiducial d_{iso} =100 kpc model is five times higher than the fiducial HMT scheme. Thus a census of AGNs at high redshifts ($z \gtrsim 7$) can distinguish between these models. Since the number density of sources in the HMT scheme is quite low (effectively 0 for redshifts $\gtrsim 8$ or 9), finding even a handful of sources at these redshifts can put stringent constrains on this seeding scheme. In Table 2, we show the number of seeds in the field for an extended redshift range by averaging from multiple random realizations of the light cone and by integrating the number density over the field volume. Almost all the averages in the redshift bins from the light cone are within 1σ of the analytically calculated value from the number density. The analytic numbers also show the drastic difference in the number of sources in the different seeding schemes at high redshifts.
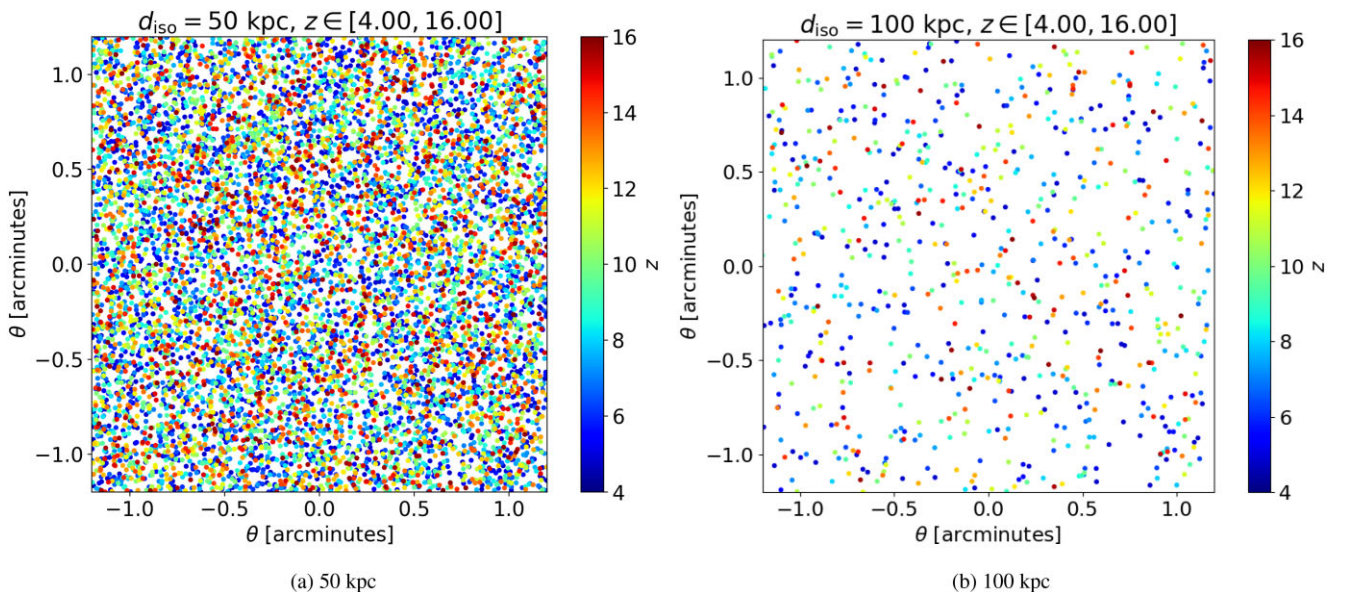

(a) 50 kpc

(b) 100 kpc

Figure 10. Synthetic HUDF consisting of only the seeded haloes for $d_{iso} = 50$ and 100 kpc cases over a redshift range from 4 to 16.
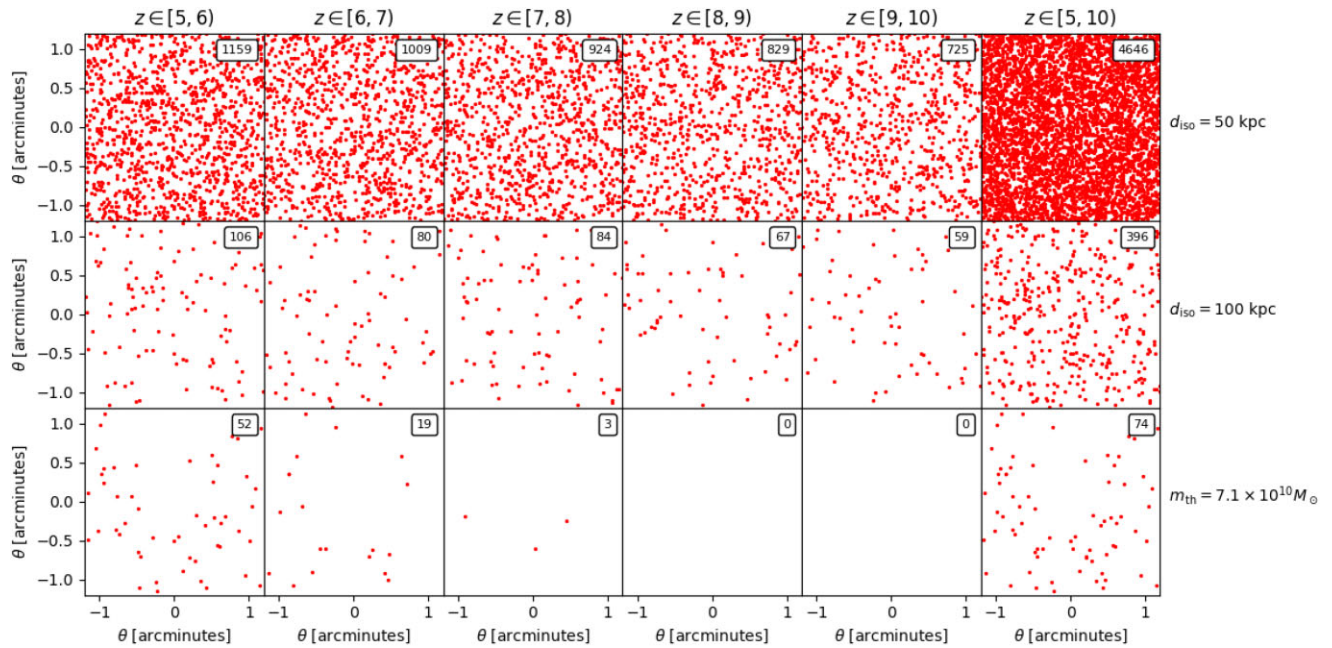
Figure 11. The distribution of SMBHs in redshift intervals in the range $z = 5$–10 in a synthetic HUDF, where the last column shows all the sources. The first row shows the case for $d_{iso} = 50$ kpc. The second row shows the case for $d_{iso} = 100$ kpc. The third row shows the distribution from the fiducial HMT scheme with $m_{th} = 7.1 \times 10^{10}$ M$_\odot$. The total number of SMBHs in each panel is indicated in the top right corners of each.

Table 2. Number of SMBHs in our synthetic HUDF, calculated by averaging over 100 random realizations of the light cone (*From light cone* column) and by integrating the global number density (*From number density* column) over the redshift ranges, for $d_{iso} = 100$ kpc and the fiducial HMT scheme with $m_{th} = 7.1 \times 10^{10}$ M$_\odot$. The errors on the averaged values correspond to 1σ deviations. Note that all the numbers are rounded to the nearest integer.

z range	100 kpc		HMT	
	From light cone	From number density	From light cone	From number density
4–5	110 ± 8	101	86 ± 19	105
5–6	92 ± 6	90	36 ± 10	49
6–7	85 ± 5	81	13 ± 6	18
7–8	74 ± 6	73	3 ± 2	7
8–9	69 ± 5	66	1 ± 1	1
9–10	60 ± 5	60	0	0
10–11	57 ± 5	54	0	0
11–12	50 ± 5	50	0	0
12–13	47 ± 4	46	0	0
13–14	42 ± 4	43	0	0
14–15	40 ± 5	40	0	0
15–16	40 ± 5	37	0	0

4 CONCLUSIONS

We have explored the implication of the Pop III.1 seeding model for cosmological distributions of SMBHs. This is a model that forms all SMBHs with a single mechanism based on the change of protostellar structure in some Pop III stars due to WIMP dark matter particle self annihilation. This leads to reduced ionizing feedback from the protostar and efficient accretion of the baryonic content of the minihalo, thus naturally leading to a characteristic seed mass of $\sim 10^5$ M$_\odot$. The model requires the Pop III.1 minihalo to form in relative isolation from other sources. Thus the Pop III.1 seeding model involves all SMBHs forming very early in the Universe, i.e. by $z \sim 25$, and with a relatively unclustered initial distribution. Indeed, compared to all other astrophysical models for SMBH formation, the Pop III.1 model involves the earliest and least clustered distribution of seeds. This implies that in the Pop III.1 model, black holes have

plenty of time to grow via accretion to explain the known high redshift quasars, without the need of sustained super-Eddington accretion.

The Pop III.1 model, while being a physical model for the formation of the whole SMBH population, is relatively simple, i.e. with only one free parameter, the isolation distance d_{iso}. This means that the model can be easily explored in cosmological volume simulations that resolve minihaloes, as was done first in Paper I. The constraint of matching an estimate for the local comoving number density of SMBHs, gives quite tight constraints on $d_{iso} \simeq 100$ kpc (proper distance). This implies most SMBHs formed at $z \simeq 30$, when the isolation distance corresponded to a comoving scale of ~ 3 Mpc. Following on from Paper I, we have explored the implications of the Pop III.1 SMBH seeding model down to low redshifts, i.e. all the way to $z = 0$, which is important to allow connection to observations, including the HUDF and local galaxy and SMBH populations. We

have also compared this model with another simple seeding scheme, i.e. the HMT model, that is commonly implemented in cosmological volume simulations.

As presented before, all SMBHs form very early in the universe, and their number density then remains approximately constant after a redshift of ~ 25. Only a small fraction of the seeded haloes merge with each other by $z = 0$. The evolution of the occupation fraction of seeded haloes shows a rise to unity for the most massive haloes by $z = 0$. However, at intermediate redshifts, there can be significant fractions of most massive haloes that are unseeded.

Our clustering analysis found that Pop III.1 seeded haloes show lower levels of clustering compared to random haloes with the same mass and number distribution as the seeded haloes, at all redshifts. However, to connect this result to observations of AGN (e.g. Allevato et al. 2014; He et al. 2018) requires development of a SMBH growth model, which is planned for a future paper in this series. We also noticed a dip in the clustering of the seeded haloes at scales smaller than the isolation distance at the mean formation redshift, which is due to the feedback suppression of the isolation bubbles. This was first discussed at $z = 10$ in Paper I, and we have shown that this suppression persists even at lower redshift, discernible down to $z \approx 1$–2.

To compare the clustering of our seeded haloes with observational data of galaxies, we turned to the galaxy clustering results from Zehavi et al. (2011). We were able to conclude that the clustering of the seeded haloes for 50 and 100 kpc isolation distances are in agreement with the observations after applying appropriate mass cuts on the halo masses. For comparison with SMBH populations, rather than just galaxies in general, it is clear that having the most complete census in a well-defined, relatively large local volume is highly desirable. Information on this local population is needed both for the total number density and to carry out a clustering analysis of the SMBHs that can be compared to the results of our models. However, obtaining such a census is challenging, given the difficulty of detecting relatively low-mass and faint SMBHs (see e.g. Reines & Comastri 2016). Another promising avenue to be explored relates to the properties of binary AGN and resulting mergers that produce gravitational waves, i.e. sensitive to the extreme end of the clustering signal. These aspects will be considered in detail in forthcoming papers in this series.

Finally, we discussed the potential of using high redshift AGN number counts in the HUDF (or other deep fields) to differentiate among seeding mechanisms and for constraining the value of isolation distance. Detection of just a small number of SMBHs at $z \gtrsim 8$ would begin to discriminate between the fiducial HMT scheme and the Pop III.1 model.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abel T., Bryan G. L., Norman M. L., 2002, Science, 295, 93
Allevato V. et al., 2014, ApJ, 796, 4
Banik N., Tan J. C., Monaco P., 2019, MNRAS, 483, 3592
Barber C., Schaye J., Bower R. G., Crain R. A., Schaller M., Theuns T., 2016, MNRAS, 460, 1147
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
Bertocco S., Goz D., Tornatore L., Ragagnin A., Maggio G., Gasparo F., Vuerli C., 2020, in Astronomical Society of the Pacific Conference Series Vol. 527, Astronomical Data Analysis Software and Systems XXIX. p. 303, preprint (arXiv:1912.05340)
Bhowmick A. K., Blecha L., Torrey P., Kelley L. Z., Vogelsberger M., Nelson D., Weinberger R., Hernquist L., 2022a, MNRAS, 510, 177
Bhowmick A. K. et al., 2022b, MNRAS, 516, 138
Boekholt T. C. N., Schleicher D. R. G., Fellhauer M., Klessen R. S., Reinoso B., Stutz A. M., Haemmerlé L., 2018, MNRAS, 476, 366
Bromm V., Loeb A., 2003, ApJ, 596, 34
Bromm V., Coppi P. S., Larson R. B., 2002, ApJ, 564, 23
Chon S., Omukai K., 2020, MNRAS, 494, 2851
Chon S., Hirano S., Hosokawa T., Yoshida N., 2016, ApJ, 832, 134
Comparat J., Prada F., Yepes G., Klypin A., 2017, MNRAS, 469, 4157
Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353
Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33
Diemer B., 2018, ApJS, 239, 35
Ebisuzaki T., 2003, in IAU Symposium Vol. 208, Astrophysical Supercomputing using Particle Simulations, P. 157
Ellis R. S. et al., 2013, ApJ, 763, L7
Feng W.-X., Yu H.-B., Zhong Y.-M., 2021, ApJ, 914, L26
Freese K., Ilie C., Spolyar D., Valluri M., Bodenheimer P., 2010, ApJ, 716, 1397
Graham A. W., 2016, in Galactic Bulges Vol. 418 Astrophysics and Space Science Library, Springer International Publishing ,Cham , p.263, preprint (arXiv:1501.02937)
Greif T. H., Bromm V., 2006, MNRAS, 373, 128
Harikane Y. et al., 2023, arXiv: preprint (arXiv:2303.11946)
He W. et al., 2018, PASJ, 70, S33
Hosokawa T., Omukai K., Yoshida N., Yorke H. W., 2011, Science, 334, 1250
Jeon J., Liu B., Bromm V., Finkelstein S. L., 2023, MNRAS, 524, 176,
Kohri K., Sekiguchi T., Wang S., 2022, Phys. Rev. D, 106, 043539
Kroupa P., Subr L., Jerabkova T., Wang L., 2020, MNRAS, 498, 5652
Lodato G., Natarajan P., 2006, MNRAS, 371, 1813
Lusso E., Valiante R., Vito F., 2023, in Bambi C., Santangelo A., eds, Handbook of X-ray and Gamma-ray Astrophysics.
Maio U., Borgani S., Ciardi B., Petkova M., 2019, Publ. Astron. Soc. Aust., 36, e020
McKee C. F., Tan J. C., 2008, ApJ, 681, 771
Monaco P., Theuns T., Taffoni G., 2002, MNRAS, 331, 587
Montero P. J., Janka H.-T., Müller E., 2012, ApJ, 749, 37
Moutarde F., Alimi J. M., Bouchet F. R., Pellat R., Ramani A., 1991, ApJ, 382, 377
Munari E., Monaco P., Sefusatti E., Castorina E., Mohammad F. G., Anselmi S., Borgani S., 2017, MNRAS, 465, 4658
Nakajima K., Ouchi M., Isobe Y., Harikane Y., Zhang Y., Ono Y., Umeda H., Oguri M., 2023, preprint (arXiv:2301.12825)
Natarajan A., Tan J. C., O'Shea B. W., 2009, ApJ, 692, 574
Norberg P. et al., 2002, MNRAS, 336, 907
O'Shea B. W., Abel T., Whalen D., Norman M. L., 2005, ApJ, 628, L5
Planck Collaboration, 2020, Planck 2018 results - VI. Cosmological parameters A&A, 641, A6
Portegies Zwart S. F., Baumgardt H., Hut P., Makino J., McMillan S. L. W., 2004, Nature, 428, 724
Rees M. J., 1978, The Observatory, 98, 210

Regan J. A., Wise J. H., Woods T. E., Downes T. P., O'Shea B. W., Norman M. L., 2020, Open J. Astrophys., 3, 15

Reines A. E., Comastri A., 2016, Publ. Astron. Soc. Aust., 33, e054

Rindler-Daller T., Montgomery M. H., Freese K., Winget D. E., Paxton B., 2015, ApJ, 799, 210

Shang C., Bryan G. L., Haiman Z., 2010, MNRAS, 402, 1249

Shinohara T., He W., Matsuoka Y., Nagao T., Suyama T., Takahashi T., 2023, preprint (arXiv:2304.08153)

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877

Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575

Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022

Spolyar D., Freese K., Gondolo P., 2008, Phys. Rev. Lett., 100, 051101

Susa H., Hasegawa K., Tominaga N., 2014, ApJ, 792, 32

Taffoni G., Becciani U., Garilli B., Maggio G., Pasian F., Umana G., Smareglia R., Vitello F., 2020, in Astronomical Society of the Pacific Conference Series Vol. 527, Astronomical Data Analysis Software and Systems XXIX, p.307, preprint (arXiv:2002.01283)

Tagawa H., Haiman Z., Kocsis B., 2020, ApJ, 892, 36

Tan J. C., McKee C. F., 2004, ApJ, 603, 383

Trebitsch M. et al., 2021, A&A, 653, A154

Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121

Vogelsberger M. et al., 2014, MNRAS, 444, 1518

Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016, MNRAS, 460, 2979

Volonteri M., Habouzit M., Colpi M., 2021, Nat. Rev. Phys., 3, 732

Wang F. et al., 2021, ApJ, 907, L1

Wise J. H., Regan J. A., O'Shea B. W., Norman M. L., Downes T. P., Xu H., 2019, Nature, 566, 85

Yang J. et al., 2020, ApJ, 897, L14

York D. G. et al., 2000, AJ, 120, 1579

Zehavi I. et al., 2011, ApJ, 736, 59

APPENDIX A: MATCHING FULL- AND LOW-RESOLUTION PINOCCHIO RUNS

We run the 59.7 Mpc box at the full resolution of 4096^3 particles and at a lower resolution of 1024^3 particles. These resolutions correspond to particle masses of $1.23 \times 10^5 \, M_\odot$ and $7.87 \times 10^6 \, M_\odot$. The minimum mass for haloes has been set to 10 particles in both cases. Fig. A1 shows the mass function of the full-resolution box at high redshift, where it is evident that the early growth of massive haloes is slower than in a universal model [in this case the fit to the friends-of-friends halo mass function of Crocce et al. (2010)]. We stress that there is no reason to believe that this analytic fit is
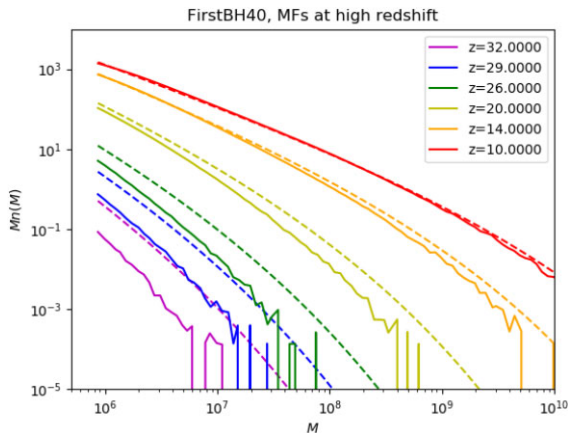
Figure A1. Halo mass function of the full-resolution box at high redshift. Lines are colour-coded in redshift (see legend). Solid lines refer to PINOCCHIO catalogs; dashed lines refer to the Crocce et al. (2010) analytic fit.
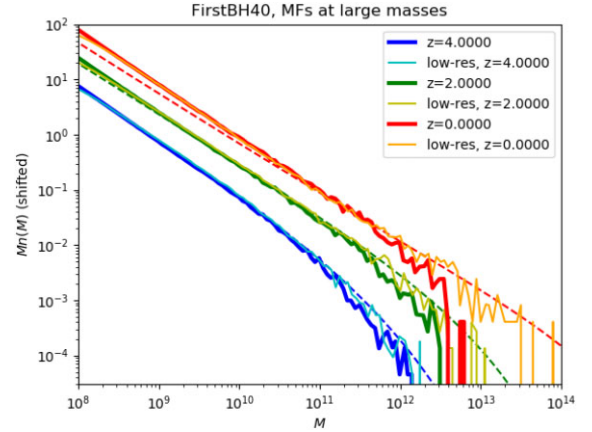
Figure A2. Halo mass function of the full- (thick solid lines) and low-resolution (thin solid lines) boxes at low redshift. Lines are colour-coded in redshift (see legend). Dashed lines are the Crocce et al. (2010) analytic fit.
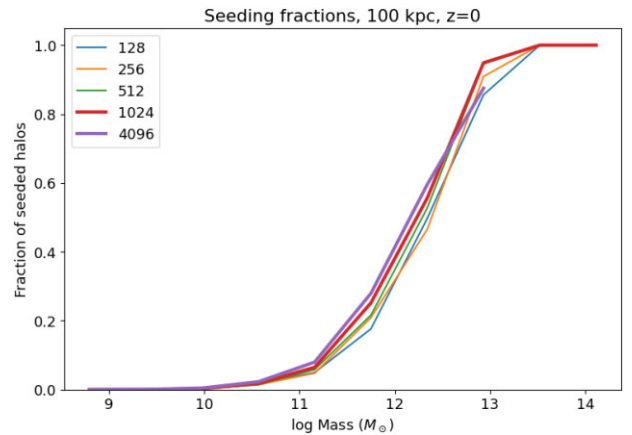
Figure A3. Fraction of haloes of a given mass that contain a seed SMBH, for $d_{iso} = 100$ kpc. Resolution is colour-coded (see legend). Thicker lines emphasize the full-resolution (4096) and low-resolution (1024) runs.

accurate at such low masses, but we conservatively assume that the disagreement is due to an inaccuracy of PINOCCHIO.

Fig. A2 shows the halo mass function for the low-resolution box (thin lines) and the full-resolution run (thick lines). At high masses, the agreement of the high-resolution box with the analytic prediction is poor while this is not the case for the low-resolution run where the box has not been divided into different domains. Fig. A3 shows the consistency of the seeding fraction among the high-resolution box and a set of lower and lower resolution runs, where seeding of haloes is decided by checking which particle in Lagrangian space contains the haloes that is seeded in the full resolution box.

APPENDIX B: LARGE-SCALE CLUSTERING MODES

When we use the estimators such as the CORRFUNC library to find the auto correlation of haloes in our 59.7 Mpc box, the correlation function only contains the clustering modes smaller than the box size. If we want to make a simplistic comparison of our results with a large survey that sampled a much larger volume, we can do so by analytically adding the larger scale clustering modes. To understand how we achieve this, we examine the analytic expression for calculating the 3D 2pcf for haloes for the entire volume of the
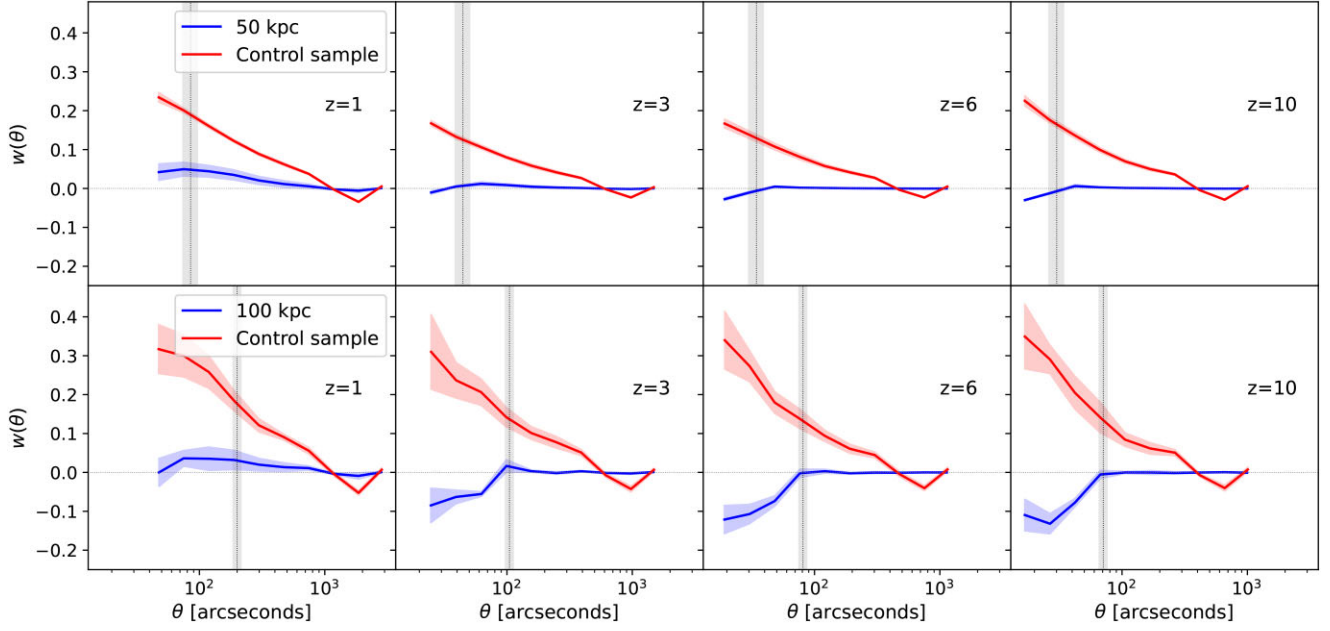
Figure B1. Evolution of angular correlation function for 50 and 100 kpc isolation distances. The large-scale modes are not added in the evaluation of this function. The labels are the same as in Fig. 8.

Universe:

$$\xi_{hh}(r) = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 b_h^2 P(k) \frac{\sin kr}{kr}, \tag{B1}$$

where ξ_{hh} is the correlation function of haloes, b_h is the halo bias, and $P(k)$ is the matter power spectrum. This integral can be split in two parts:

$$\xi_{hh}(r) = \underbrace{\frac{1}{2\pi^2} \int_0^{k_{\mathrm{box}}} dk\, k^2 b_h^2 P(k) \frac{\sin kr}{kr}}_{\text{Large scale contribution } \xi_{\mathrm{LS}}(r)}$$
$$+ \underbrace{\frac{1}{2\pi^2} \int_{k_{\mathrm{box}}}^\infty dk\, k^2 b_h^2 P(k) \frac{\sin kr}{kr}}_{\text{PINOCCHIO contribution } \xi_{\mathrm{PIN}}(r)} \tag{B2}$$
$$= \xi_{\mathrm{LS}}(r) + \xi_{\mathrm{PIN}}(r),$$

where $k_{\mathrm{box}} = 2\pi/L_{\mathrm{box}}$, with $L_{\mathrm{box}} = 59.7$ Mpc in our box. The large-scale contribution refers to the clustering modes of radial scale going

from L_{box} to infinity, and the PINOCCHIO contribution refers to all the modes of radial scale from 0 to L_{box}. Since the correlation estimator returns ξ_{PIN}, we calculated the large-scale contribution by using the linear matter power spectrum from CAMB python library and halo bias from COLOSSUS python library (Diemer 2018), using the bias model of Comparat et al. (2017), and then numerically integrated the power spectrum to obtain ξ_{LS}.

To make a direct continuation of the angular clustering as shown in Banik et al. (2019; their fig. 10), we present the angular clustering evolution of seeded haloes in Fig. B1 without the large-scale corrections added. This figure and Fig. 8 essentially show the same information, with the only difference that the figure presented here is in angular scale, and without the large-scale modes.

This paper has been typeset from a TEX/LATEX file prepared by the author.