Building A Semi-Analytic Black Hole Seeding Model Using IllustrisTNG Host Galaxies

Analis Eolyn Evans[1⋆], Laura Blecha [1], Aklant Kumar Bhowmick [1]
[1] *Department of Physics, University of Florida, Gainesville, FL 32611, USA*

11 December 2024

ABSTRACT

A major open question in astrophysics is the mechanisms by which massive black holes (BHs) form in the early Universe, which pose constraints on seeding models. We study BH formation and evolution in a flexible model combining the cosmological IllustrisTNG (TNG) simulations with semi-analytic modeling in post-processing. We identify our TNG model hosts based on various criteria including a minimum gas mass of 10^7-10^9 $M_\odot$, total host mass of $10^{8.5}$-$10^{10.5}$ $M_\odot$, and a maximum gas metallicity of 0.01 - 0.1 $Z_\odot$. Each potential host is assigned a BH seed with a probability of 0.01 - 1. The populations follow the TNG galaxy merger tree. This approach improves upon the predictive power of the simple TNG BH seeding prescription, narrowing down plausible seeding parameter spaces, and it is readily adaptable to other cosmological simulations. Several model realizations predict $z \lesssim 4$ BH mass densities that are consistent with empirical data as well as the TNG BHs. However, high-redshift BH number densities can differ by factors of ~ 10 to $\gtrsim 100$ between seeding parameters. In most model realizations, $\lesssim 10^5$ $M_\odot$ BHs substantially outnumber heavier BHs at high redshifts. Mergers between such BHs are prime targets for gravitational-wave detection with LISA. The $z = 0$ BH mass densities in most realizations of the model agree well with observations, but our strictest seeding criteria fail at high redshift. Our findings strongly motivate the need for better empirical constraints on high-z BHs, and they underscore the significance of recent AGN discoveries with JWST.

Key words: black holes: general, galaxies: groups: general

1 INTRODUCTION

Supermassive black holes (SMBHs) are ubiquitous in the centers of massive galaxies and are key drivers of galaxy evolution. Observations of luminous active galactic nuclei (AGN) at $z \sim 6 - 11$ (e.g., Fan et al. 2006; Mortlock et al. 2011; Bañados et al. 2018; Larson et al. 2023) indicate that the earliest massive BHs assembled in less than $\sim 0.5 - 1$ Gyr of cosmological time after the Big Bang. Many of the highest redshift AGN candidates have been discovered only recently with JWST and have lower inferred masses than the extreme $\sim 10^9$ $M_\odot$ quasar population that previously provided our only constraints on BH populations at $z \sim 6 - 7.5$ (e.g., Onoue et al. 2023; Maiolino et al. 2023; Übler et al. 2023; Juodžbalis et al. 2023). These exciting discoveries are placing new constraints on the prevalence and masses of accreting BHs at early times, while simultaneously raising new questions about the interplay between BH formation and growth in the early Universe.

A variety of different theoretical scenarios for BH formation (or "seeding") have been proposed, which could produce seeds at a wide range of masses and redshifts. For example, the earliest Population III (Pop III) stars were massive and essentially metal-free, and they should have therefore created massive BH remnants. Population III stars form in the gravitational potential well of dark matter minihalos that collapse at $z \sim 20$ due to high matter density fluctuations; these are expected to form $\approx 10^2 - 10^3$ $M_\odot$ BH seeds (Volonteri et al. 2003). However, such seeds would require sustained periods of super-Eddington accretion to reach the supermassive regime by the epoch of the earliest quasars, and they would still require reasonably efficient growth to produce less extreme high-redshift AGN, such as those recently discovered by JWST.

A possible solution to these tight constraints on BH growth timescales is that seeds form at higher initial masses. One promising scenario is that of direct collapse BHs (DCBHs), seeded when a massive, metal-free gas cloud collapses directly into a BH or supermassive star (SMS) with a mass of $\approx 10^5$ $M_\odot$ (e.g., Rees 1984; Bromm & Loeb 2003; Begelman et al. 2006; Latif & Ferrara 2016; Inayoshi et al. 2020). This monolithic collapse requires that gas fragmentation be prevented by some means. One possibility is the dissociation of molecular hydrogen through UV radiation in the Lyman-Werner band, which could be provided by nearby star-forming regions. Lyman-Werner radiation prevents the fragmentation

⋆ analislawrence@ufl.edu

that would normally happen in low-temperature gas near the cosmological Jeans mass (Bromm & Loeb 2003). Alternately, dynamical heating by mergers or turbulent cold flows may suffice to suppress fragmentation and allow DCBH formation (e.g., Mayer et al. 2010, 2015; Wise et al. 2019; Latif et al. 2022; Zwick et al. 2023). Magnetic fields have been proposed as another means to catalyze and trigger the formation and early growth of massive BH seeds, by suppressing fragmentation and star formation and boosting the accretion flow to newly formed DCBH seeds (Begelman & Silk 2023). These various possible scenarios for DCBH formation vary widely in their predicted ranges of seed formation redshifts (anywhere from $z \sim 6-20$) and masses (anywhere from $\sim 10^4 - 10^8\ M_\odot$) (e.g., Bromm & Loeb 2003; Lodato & Natarajan 2006; Agarwal et al. 2012; Mayer et al. 2024). Any mergers between heavy DCBH seeds would create an additional avenue for growth, and these would be prime candidates for gravitational-wave (GW) detection with the Laser Interferometer Space Antenna (LISA) (Amaro-Seoane et al. 2017, 2023).

Additionally, SMBH seeds can naturally be formed via successive mergers of massive stars or stellar remnants, which could form intermediate-mass BHs (IMBHs) of $\approx 10^3 - 10^5\ M_\odot$ (e.g., Bond et al. 1984; Madau & Rees 2001; Bromm & Larson 2004; Davies et al. 2011). In the stellar collision scenario, a runaway process ensues as the collisional cross section grows, creating a supermassive star that can collapse to an IMBH (e.g., Omukai et al. 2008; Yajima & Khochfar 2016; Sakurai et al. 2017; Boekholt et al. 2018; Tagawa et al. 2020). Alternately, an IMBH could form via successive mergers between stellar-mass BHs. This channel would likely be aided by the presence of copious gas within the star cluster to prevent BH ejections, which could also provide fuel for an inital burst of BH growth shortly after formation (e.g. Davies et al. 2011; Ryu et al. 2016; Tagawa et al. 2016). In the nearby Universe, many BHs co-exist with dense, massive regions of stars near galactic centers known as nuclear star clusters (NSCs), kindling the idea that NSCs may have come before the SMBH (Neumayer et al. 2020; Askar et al. 2022, and references therein). The formation pathways of IMBHs within NSCs depend on the their mass, density, and spin (Miller & Hamilton 2002; Greene et al. 2020; Fragione & Silk 2020). Merger events in nuclear star clusters are potential GW sources for LISA as well as for ground-based GW detectors such as the LIGO-Virgo-KAGRA (LVK) collaboration (Abbott et al. 2016; Jiang & Huang 2022).

Not much is understood about early-Universe BH-galaxy co-evolution due to observational limitations. For example, selection biases make it unclear how the BH mass - stellar mass relation may evolve above $z \sim 2$ (Shields et al. 2003; Jahnke et al. 2009; Suh et al. 2020). One key observational bias is that it is difficult to observe faint quasars especially at high redshift, which could reveal more about the entire BH population than bright quasars (Habouzit et al. 2022). Accordingly, AGN observations show increasing uncertainties in the BH mass function (BHMF) up to $z \sim 6$ (Merloni & Heinz 2008; Shankar et al. 2009; Cao 2010). The faint, early-Universe quasars can help explain the stepping stones for assembly of the massive, most luminous ones through constraining early demographics. JWST can detect rest-frame UV and optical light from faint quasars that has previously been inaccessible for high-redshift quasars (e.g., Decarli et al. 2012; Marshall et al. 2020). There have already been numerous high-redshift BH candidates identified in JWST data, including low-mass candidates (e.g., Harikane et al. 2023; Matthee et al. 2023; Labbe et al. 2023; Juodžbalis et al. 2023). This trove of early discoveries is a promising indication that JWST will continue to reveal a great deal about early BH-galaxy co-evolution and the epoch of reionization.

As noted above, BH assembly mechanisms are also of great interest owing to the potential for GW detections of BH mergers with LISA and LVK, as well as next-generation ground-based GW detectors. LISA will be revolutionary for our understanding of BH assembly in a regime where electromagnetic (EM) constraints are sparse or non-existent, with the capacity to detect BH mergers in the mass range $\sim 10^4 - 10^7 M_\odot/(1+z)$ out to $z \sim 20$ (Vecchio et al. 2004; Lang & Hughes 2006, 2007; Amaro-Seoane et al. 2017, 2023).

Pulsar Timing Array (PTA) experiments are sensitive to GWs in the $\lesssim$ nanoHertz - microHertz range, corresponding to $\sim 10^9 M_\odot$ BH binaries. Recently, PTAs around the globe presented strong evidence for a stochastic GW background that is consistent with the expected signal from a cosmological population of BH binaries (Agazie et al. 2023a; Antoniadis et al. 2023; Reardon et al. 2023; Xu et al. 2023). Future PTA data will constrain the spectral shape of this background and the (an)isotropy of its origin on the sky, both of which will provide key insight into SMBH binary evolution.

Improved predictions of GW and EM signatures from different BH assembly channels are needed to interpret data from the upcoming observations described above. Many theoretical studies of BH formation and growth rely on semi-analytic models (SAMs), which have the unique ability to probe a wide range of seeding scenarios with little computational expense (e.g., Sesana et al. 2007; Volonteri & Natarajan 2009; Barausse 2012; Valiante et al. 2018; Ricarte & Natarajan 2018; Sassano et al. 2021).

Most of these SAMs have thus far relied on tracking BH seeding and growth over halo merger trees constructed using analytic formulations such as the Press-Schechter (Press & Schechter 1974) or dark-matter-only cosmological simulations. However, by construction, SAMs cannot trace the detailed hydrodynamics of the gas or the internal structure of galaxies. This poses as a significant limitation on modeling BH seed formation, which crucially relies on the local gas conditions within halos.

Alternatively, BH evolution can also be modeled in cosmological hydrodynamics simulations, which (unlike SAMs) do solve the gas hydrodynamics along with sub-grid prescriptions for BH seeding, accretion, and feedback. Numerous large-volume cosmological simulations including Illustris, IllustrisTNG (hereafter TNG), SIMBA, EAGLE, and Horizon-AGN have been shown to produce results consistent with many observed properties of galaxy and BH populations, including the BH-bulge relation (Vogelsberger et al. 2014; Dubois et al. 2014; Schaye et al. 2015; Weinberger et al. 2017; Pillepich et al. 2018b; Davé et al. 2019). However, these simulations have a major drawback compared to SAMs, in that their huge computational expense prohibits exploring large parameter spaces. Further, most of these simulations still cannot directly resolve low-mass BH seeds.

Due to these challenges, most large-volume cosmological simulations adopt very simplistic seed models. For example, many simulations seed $\sim 10^5 - 10^6\ M_\odot$ BHs in halos above

a fixed mass threshold of $\sim 10^9 - 10^{10}\ M_\odot/h$ (Vogelsberger et al. 2014; Khandai et al. 2015; Schaye et al. 2015; Feng et al. 2016; Nelson et al. 2019a). While these simple prescriptions reproduce local BH populations reasonably well, their predictive ability at high redshift is limited, and they cannot distinguish between different BH seeding channels. Several simulations have also seeded BHs based on local gas properties in cosmological simulations (e.g., Taylor & Kobayashi 2014; Tremmel et al. 2017; Wang et al. 2019). In particular, they create BH seeds from gas cells that exceed a critical density threshold while remaining metal-poor. These prescriptions are much more representative of theoretical seeding channels such as Pop III, NSC and DCBH seeds, all of which are expected to form exclusively in regions of dense and metal-poor gas. However, at coarse gas mass resolutions, the poor resolution convergence of gas properties could impact of resolution convergence of the final BH populations.

At very high gas mass resolutions ($\sim 10^3 - 10^4\ M_\odot$) typical of zoom simulations, gas-based seed models do start producing reasonably well converged BH populations (Bhowmick et al. 2021). Zoom simulations are also relatively computationally inexpensive, which has allowed several recent works to explore a wide range of gas-based seed models (Bhowmick et al. 2021, 2022a,b). However, since zoom simulations typically focus on a small biased region of the universe, they cannot be readily compared to observations.

In this work, we adopt a new approach that harnesses the strengths of conventional SAMs and full hydrodynamics simulations, while mitigating the limitations inherent in each approach. We develop a novel, SAM-based BH seed modeling approach that can trace BH evolution across merger trees within any existing cosmological hydrodynamical simulation. By doing this, our seeding prescriptions can be informed by the detailed gas properties of halos on the merger trees, which are inaccessible in conventional SAMs. A similar approach was taken by DeGraf & Sijacki (2020) wherein BH growth histories within Illustris were reconstructed for subsets of the simulated BH population. These subsets were selected by introducing additional seeding criteria beyond the default seed model used by Illustris, such as spin and metallicity based seeding. They found that the total BH merger rate can be substantially impacted by the introduction of these seeding criteria. In contrast to DeGraf & Sijacki (2020), our model places new seed BHs in the subhalo merger trees that are completely independent of the BHs that formed during the actual run of the parent simulation. This enables us to study a wide variety of BH formation model realizations, including criteria that are more lenient than those used on-the-fly during the simulation run.

For our parent simulation, we use the highest-resolution run of the TNG suite, TNG50-1. In Appendix B, we use the lower-resolution versions of TNG50 for convergence tests. Unless otherwise specified, "TNG50" refers to the highest-resolution TNG50-1 simulation in the remainder of the paper. In this initial study, we present our modeling framework and initial results for the highest resolution simulation in the TNG suite, leaving its application to larger-volume simulations for future work. Although the TNG50 volume is not large enough to capture rare, extreme objects like the population of $\sim 10^9\ M_\odot$ $z \sim 6$ quasars, it well reproduces the properties of galaxies across a wide range of masses, morphologies, and cosmic epochs (e.g., Pillepich et al. 2018b; Springel et al. 2018). With a gas mass resolution of $8 \times 10^4\ M_\odot$, TNG50 offers a resolution comparable to zoom simulations over a reasonably large volume of $(50\ \mathrm{Mpc})^3$ (Nelson et al. 2019b). This allows us to use well resolved gas properties to explore seeding criteria and examine their level of convergence with resolution, enabling construction of a model that can be readily applied to larger-volume, lower-resolution simulations. In service of this goal, we also choose to focus on halo-averaged properties rather than basing seed formation on the properties of individual gas cells. The model assumptions include allowing a maximum of one BH per massive, low-metallicity, galaxy or galaxy group with a sufficient gas reservoir. A key advantage is that the stellar, gas, and host properties that inform the seed model are directly attainable from the simulation. This avoids the use of an empirical framework to derive the baryonic properties, which is commonly used in most SAMs. Our seeding criteria are motivated by theoretical seed formation scenarios, in that a range of proposed BH seeding channels require a minimum halo mass as well as a sufficient quantity of low-metallicity gas. This includes direct collapse formation scenarios as well as intermediate-mass seed formation in stellar clusters. As noted above, the ranges of BH seed masses span orders of magnitude and have some overlap between massive and intermediate-mass BH formation channels. Thus, rather than tying specific sets of model parameters to particular theoretical seeding channels, we carry out a systematic study of the allowed parameter combinations, when compared to observations.

This paper is organized as follows. Section 2 summarizes key features of the IllustrisTNG simulations, describes our methodology for constructing a TNG-based SAM for BH seeding and growth, and details the parameter space explored in this work. Section 3 present our results, including an analysis of the properties of high-z TNG halos (Section 3.1), a verification that our SAM can successfully reproduce the TNG BH population (Section 3.2), and a detailed analysis of the BH populations produced by our SAM, including BH number and mass density evolution and local BHMFs (Sections 3.3 & 3.4). We summarize and conclude in Section 4. Throughout this paper, we assume the same cosmology as the TNG simulation suite (as specified below).

2 METHODS

2.1 IllustrisTNG simulations

The TNG simulation project is a cosmological magnetohydrodynamical simulation suite (Marinacci et al. 2018; Pillepich et al. 2018b; Springel et al. 2018; Naiman et al. 2018). The initial cosmological conditions are $\Omega_{\Lambda,0} = 0.6911$, $\Omega_{m,0} = 0.3089$, $\Omega_{b,0} = 0.0486$, $\sigma_8 = 0.8159$, $n_s = 0.9667$, and $h = 0.6774$, taken from Planck collaboration observations of the cosmic microwave background (Ade et al. 2016). These simulations were carried out with the quasi-Lagrangian AREPO code (Springel 2010; Pakmor et al. 2011; Pakmor & Springel 2013; Weinberger et al. 2020) in which gravitational equations are coupled with magnetohydrodynamics (MHD) equations. The gravity is solved using a tree-particle-mesh N-body algorithm, and the MHD is solved using an adaptive unstructured mesh that is constructed by performing a Voronoi tesselation of the simulation volume.

AREPO implements sub-grid modeling for a variety of physical processes that cannot be directly resolved in current cosmological simulations. These include gas cooling, star formation and evolution, chemical enrichment and feedback. Star formation happens within gas above a critical threshold density of 0.1 cm^{-3} (Hernquist & Springel 2003). Stellar evolution assumes an initial mass function from Chabrier (2003). The stellar feedback includes energy released from AGB stars and supernovae, and it is primarily responsible for depositing metals on to the surrounding gas. Further details about the implementation of these processes are described in Pillepich et al. (2018a). The BH-related sub-grid physics models will be discussed in more detail below and in Sections 2.1.1 and 2.1.2.

Subhalo and halo catalogs are saved for each snapshot with a wide range of quantities including gas-phase metallicities, star-formation rates, stellar, BH, and total host masses, velocity dispersion, and the number of BHs per subhalo or halo. The Friends-of-Friends (FoF) algorithm (Press & Davis 1982; Huchra & Geller 1982; Merchán & Zandivarez 2005) groups DM particles together if they are within 0.2 times the mean separation (van Daalen & Schaye 2015). The halos can be generally identified as groups of galaxies. The subhalo catalog is computed using `SUBFIND` (Springel et al. 2001); subhalos can generally be identified as galaxies in the simulation. For a negligible number of catalog objects near the resolution limit, the algorithm cannot distinguish galaxies versus spurious clumps; these are excluded from our analysis based on their tendency to have very low masses.

TNG has overall produced good agreement for BH and galaxy properties, including, but not limited to, BH scaling relations (Li et al. 2020), correlations between SMBH mass and X-ray temperature of the hot gaseous halos pervading host galaxies, the underlying SMBH-halo mass relation (Truong et al. 2021), the BH-stellar bulge mass relation (Weinberger et al. 2017; Habouzit et al. 2021), and anisotropic black hole feedback causing quiescent satellites to be found less frequently along the minor axis of their central galaxies (Martín-Navarro et al. 2021). Our primary simulation TNG50-1 has a $(50\ \mathrm{Mpc})^3$ box that includes 2160^3 gas cells (Nelson et al. 2019b).

2.1.1 BH formation and evolution

Seeding, growth, and feedback are all important processes in BH evolution. In the TNG simulation, BHs of seed mass $8 \times 10^5\ h^{-1} M_\odot$ are placed in halos with dark matter halos exceeding a total mass threshold of $5 \times 10^{10}\ h^{-1} M_\odot$ (Weinberger et al. 2017). More specifically, the densest gas particle of a halo is converted to a BH particle if the halo does not already contain a BH.

BH growth is modeled by Eddington-limited Bondi accretion (and can also be facilitated through mergers):

$$\dot{M}_{\mathrm{Edd}} = \frac{4\pi G M_{\mathrm{BH}} m_p c}{\epsilon_r \sigma_T}, \tag{1}$$

$$\dot{M}_{\mathrm{Bondi}} = \frac{4\pi G^2 M_{\mathrm{BH}}^2 \rho}{c_s^3}, \tag{2}$$

$$\dot{M}_{\mathrm{BH}} = \min(\dot{M}_{\mathrm{Bondi}}, \dot{M}_{\mathrm{Edd}}), \tag{3}$$

where M_{BH} is the BH mass, ϵ_r is the radiative efficiency (set to 0.2 in TNG), σ_T is the Thomson scattering cross-section, m_p is the proton mass, and ρ & c_s are the gas density and sound speed, respectively, in cells neighboring the BH. The feedback model for BHs in TNG assumes thermal or kinetic energy feedback modes from the AGN. The kinetic mode is comparably more efficient and is the dominant means for SMBH growth for BHs above $\approx 10^8\ M_\odot$ at low accretion rates relative to the Eddington limit (Weinberger et al. 2017). The thermal mode of AGN feedback is associated with high accretion rates and jets where, along with mergers, it is responsible for the star-formation quenching of massive galaxies (Weinberger et al. 2017).

2.1.2 Merger Trees

The Sublink merger trees (Rodriguez-Gomez et al. 2015) include a descendants tree branch with galaxy identifiers that allow merger tracking. TNG descendant selection is performed by first identifying subhalo descendant candidates, scoring them with a merit function based on its particles' binding energy rank, and deeming the descendant as the one with the highest score (Rodriguez-Gomez et al. 2015). The progenitor subhalo's constituent particles are matched with the same particles on descendant subhalos. Following TNG's descendant links in our reconstructed TNG merger trees, starting from points at which galaxies meet the model seeding criteria, we are able to follow these populations of galaxies and their BHs, each with its own unique merger history.

2.2 Simulation Analysis: Semi-Analytic Black Hole Seeding Model

2.2.1 Identifying BH seeding sites

For our novel, hybrid SAM, we apply host criteria to identify BH seeding sites within TNG in a post-processing approach. Gas mass and metallicity properties in TNG halos are examined, since the gas-based BH seeding model requires low metallicity as well as a large enough gas reservoir to form seeds. We remain agnostic as to the exact seed formation scenario, motivating our seeding criteria under the assumption that a range of seeding channels, including intermediate-mass seed formation in dense star clusters and massive seed formation via direct collapse, require low metallicity and a large enough gas reservoir to form seeds. The specific metallicity and mass requirements depend on the details of the seeding channel, which may depend on a variety of other local physical properties as well (e.g., Schneider et al. 2006; Ahn et al. 2007; Smith et al. 2009; Omukai et al. 2010; Bromm & Yoshida 2011). In particular, the onset of star formation, and the subsequent feedback from supernovae, play an important role in regulating the mass and abundance of BH seeds (e.g., Clark et al. 2008; Omukai et al. 2008; Inayoshi & Omukai 2012; Becerra et al. 2015; Ge & Wise 2017). Because one of the goals of this work is the construction of a seeding prescription that can be applied to larger-volume, lower-resolution simulations, we focus on halo-averaged quantities like average gas-phase metallicity and total gas mass rather than the properties of individual gas cells, and we determine the regime in which these quantities are well converged with resolution.

Mass-metallicity histograms from Figure 1 give insight on reasonable choices of BH seeding constraints for our model.

We define the total gas mass and metallicity of a galaxy as that within $R_{\rm max}$, the radius at which the galaxy reaches its maximum rotational velocity. To ensure that the subhalos selected for BH seeding are reasonably well resolved and contain a large enough gas cloud with the potential to collapse, we implement cuts on the minimum total and gas mass. Each model realization with a unique variation in minimum mass and maximum metallicity yields a large sample of galaxies with the potential to form BHs. The questions become: what seeding criteria produce reasonable BH populations and how do these the results of these criteria compare to the TNG population? By comparing our results with the observed BH population, we can constrain the parameter space of seeding criteria and inform future studies of BH formation and evolution. Additionally, since TNG is known to produce good agreement with well-established local BH scaling relations, it provides a useful benchmark to compare the predictions of our SAM based seed model realizations, particularly at higher redshifts wherein the empirical constraints are more uncertain.

As the first criteria for identifying potential BH seeding sites, we require the host galaxy to have a minimum total and gas mass. We implement total mass cuts ranging from $10^{8.5} - 10^{10.5}$ $M_\odot$ and gas mass cuts ranging from $10^7 - 10^9$ $M_\odot$. These values are well above the baryonic and dark matter mass resolution of TNG50, $m_{\rm b} = 8.5 \times 10^4$ $M_\odot$, ensuring that the selected galaxies are well-resolved. We also explore the requirement for seeded galaxies to have nonzero star-formation rates, but in practice, we find that nearly all TNG galaxies that meet the above mass criteria are also star-forming (see Figure 1).

We additionally require the potential seeding hosts to have low gas metallicity. The primordial metallicity set initially for several chemical species in TNG50 is a mass fraction of 10^{-10}, or $10^{-8.1} Z_\odot$. The maximum metallicity values in our BH seeding model, set to $Z_{\rm max} = 10^{-1}$, $10^{-1.5}$, or $10^{-2} Z_\odot$, are consistent with the findings of no fragmentation occurring for gas cloud metallicities up to $Z \sim 0.1$ $Z_\odot$ for number densities as high as $10^5 {\rm cm}^{-3}$, and where metal-line cooling does not happen effectively below 10^{-3} $Z_\odot$ (Jappsen et al. 2009). By choosing maximum metallicity values no lower than 10^{-2} $Z_\odot$, we also ensure that our results are well converged with resolution. (see Appendix B).

Additional, complex physical processes may be involved in the formation of a BH seed that are not captured by the above seeding criteria. To account for this possibility, we also consider probabilistic seeding with a random seeding probability $f_{\rm seed} < 1$, specifically down to $f_{\rm seed} = 0.01$. Each galaxy (subhalo) or galaxy group (halo) that meets all other seeding criteria in a given simulation snapshot has a probability $f_{\rm seed}$ of forming a BH in that snapshot.

Because we select BH seeding sites based solely on galaxy properties as they were computed during the actual TNG50 run and do not recompute the galaxy properties for our new SAM based seed model, there is an inherent inconsistency regarding the impact of BH feedback on host galaxies. Galaxies that have BHs within the TNG simulation (many of which will also contain BHs in our model) will experience AGN feedback effects, while galaxies that form BHs in our model but not in TNG will not experience any impact from AGN feedback. However, numerous theoretical and observational studies demonstrate that AGN feedback dominates over stellar feedback primarily in massive, low-redshift galaxies (e.g., Torrey et al. 2020; Fluetsch et al. 2019; Valentini et al. 2021). The primary focus of this work, in contrast, is on the formation and early growth of BHs at high redshift. Even within the high-redshift regime, massive galaxies will generally have BHs in both TNG50 and in our post-processing model. Thus, we expect this limitation to have a minimal effect on our results, and we consider this a worthwhile trade-off for the flexibility and computational efficiency of exploring a wide range of seeding model realizations based on the TNG50 galaxy populations. The high-redshift BH seeding sites in halos have not yet undergone substantial metal enrichment through star formation, so we do not impose a minimum stellar mass criterion in order to form a BH seed, except to require that the stellar mass be nonzero.

2.2.2 *Merger-tree Modeling of BH Populations*

To estimate the cosmic evolution of BH populations for each seeding model realization, we follow `SUBFIND` galaxy merger trees, assigning unique growth histories based on seeding criteria. We trace the progenitors and descendants of galaxies that satisfy the chosen seeding criteria. The `SUBFIND` merger trees are based on the evolution of subhalos, while most seeding prescriptions in cosmological simulations rely on the properties of halos. Accordingly, our fiducial seeding criteria are applied to halo properties, but to trace these seeding sites through the merger trees, we assume seeding the central subhalo (CSH) in each halo (defined to be the most massive subhalo in a given halo). Ultimately, halo identification is then performed on the unique merger trees formed by CSH proxies (of the halo). Appendix A indicates that this choice does not have a strong influence on our results. The use of CSH proxies does limit the model from seeding BHs in satellite galaxies within halos, but in practice, the population of TNG satellites that meet the model seeding criteria and also have BHs is small (see Figure A1).

Regardless, this approach does necessitate the simplifying assumption of one BH per halo, such that when two galaxies merge and each contains a BH seed, we assume the BHs also promptly merge. This treatment gives a lower limit on the BH number densities and the merger timescales for each seeding model realization. It is a rough approximation over the course of descendant evolution, because BH merger timescales can be several Gyr (e.g., Kelley et al. 2017). (Interestingly, recent analysis of the PTA evidence for a stochastic GW background suggests that short inspiral timescales are favored by the data (Agazie et al. 2023b). These early results are still too tentative to provide a robust justification for our simplifying assumption, however.) A detailed study (e.g., using different BH growth or dynamical friction models) aimed at examining BH merger rates or LISA event rates would warrant a more realistic treatment of BH binary inspiral timescales. Gravitational recoil would also be important to consider BH retention within the galaxies (e.g., Blecha & Loeb 2008). We plan to focus on these details in future studies.

2.2.3 *Modeling BH Growth*

As noted in Section 2.2.1, our post-processing scheme for seeding BHs and tracing them through galaxy merger trees

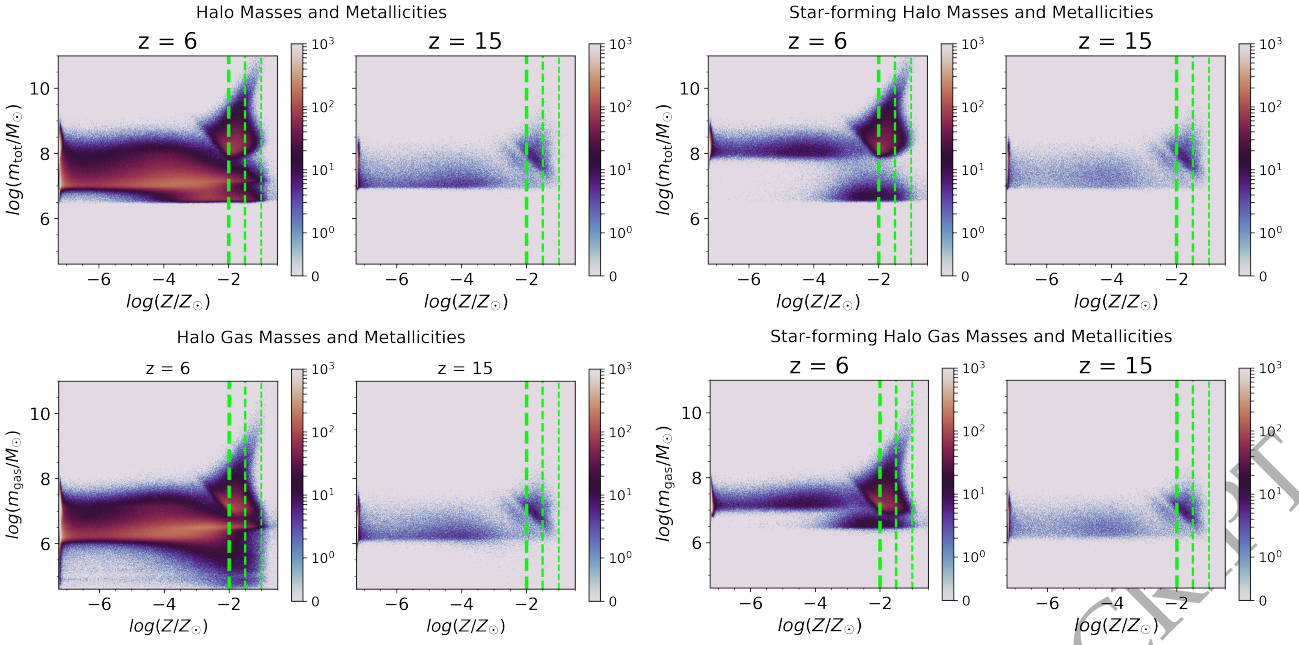

Figure 1. We use 2D histograms of gas mass and total mass versus gas metallicity to illustrate the properties of high-redshift halos; these distributions motivate our choices of seeding criteria. The maximal gas metallicity values used in our model are $Z_{\rm max}/Z_\odot = 10^{-1}$, $10^{-1.5}$, or 10^{-2} (shown by the thinnest to thickest dashed green lines, respectively). The gas properties of each halo are averaged within the total gas cells for each halo. The top-left panels are total galaxy group mass - gas metallicity histograms at $z = 6$ and 15, while the top-right panels show the same data for the subset of star-forming halos (those with SFR > 0). In the same order, the bottom panels show the distributions of gas mass (rather than total halo mass) versus gas metallicity; again, the bottom-left panels show all halos, while the bottom-right panels show only star-forming halos. In all cases, the most lenient metallicity criterion ($Z_{\rm max}/Z_\odot = 0.1$) encompasses nearly 100% of halos in each snapshot, while the strictest metallicity cut ($Z_{\rm max}/Z_\odot = 0.01$) includes only 94% of halos and 62% of star-forming halos by $z = 6$.

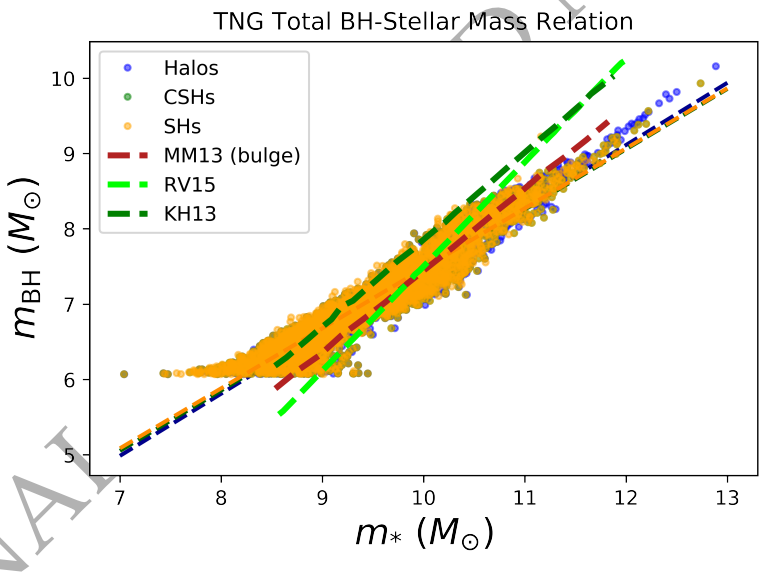

Figure 2. The $z = 0$ TNG BH mass-total stellar mass relation for TNG halos, CSHs, and subhalos is shown (blue, green, and yellow points, respectively), along the best-fit lines for each. For comparison, the empirical BH mass-total stellar mass relation for ellipticals and bulge galaxies from Reines & Volonteri (2015) (RV15) is shown in light green. We also show the BH mass-stellar bulge mass relations for ellipticals and bulge galaxies from Kormendy & Ho (2013) (KH13; dark green) and from McConnell & Ma (2013) (MM13; red). Note that the TNG relation is derived for total stellar mass in the halos for simplicity rather than performing a kinematic decomposition to derive the bulge masses. The relation is nearly identical to that of subhalos and CSHs and it is purposed for assigning BH masses in the model more similarly to how TNG does. Note also that some caution is warranted in comparing the TNG BH-total stellar mass relations with the KH13 and MM13 BH-bulge mass relations; while these are the same quantity for elliptical galaxies, they differ in galaxies that have both a bulge and a disk component.

provides BH number densities and occupation fractions, but it does not allow for BH masses and accretion rates to be obtained directly from the simulation. In order to compare our seeding model results with empirical measurements of BH mass functions and mass density evolution with redshift, we employ a simple prescription to assign masses to the BHs when they form and as they evolve through time.

We implement the $z = 0$ TNG BH mass - total stellar mass scaling relation *with scatter* on our model. Specifically, the mass of each BH is assumed to evolve with the total stellar masses of the halos. In Figure 2 we show the TNG BH-stellar mass relation within TNG halos, CSHs, and subhalos. For comparison, we also show the empirical BH mass - total stellar mass scaling relation for ellipticals and bulge galaxies from Reines & Volonteri (2015), as well as the BH mass - stellar bulge mass relations for ellipticals and bulges from Kormendy & Ho (2013) and McConnell & Ma (2013). In comparing with the scaling relations from these latter two works, it is important to keep in mind that while total stellar mass and bulge mass are the same quantity for ellipticals, this is not true of bulge galaxies that also have a disk component. Nonetheless, as these scaling relations are widely used in the literature, including for qualitative comparisons to BH mass - total stellar mass relations from the Illustris simulations (Sijacki et al. 2015), we include them here for reference. We obtain a fit for the BH stellar mass scaling relation in TNG halos and use this relation to determine BH masses in our model. Figure 2 illustrates that nearly identical fits are obtained for scaling relations using TNG subhalos and CSHs instead of halos. Below is the TNG BH mass - total stellar mass scaling relation, where $\sigma = 0.135$, $\alpha = 6.68$, $\beta = 0.79$, and M_* is the total stellar mass of the halo:

$$\log M_{\rm BH} = \alpha + \beta \log\left(\frac{M_*}{10^9 M_\odot}\right) + \log \mathrm{N}(\mu, \sigma^2) \qquad (4)$$

We obtained σ using the same method in (McConnell & Ma 2013), where $\chi^2 = 1 = N_{dof}$ and μ is zero-centered (see their equation 4). Since the BHs are assumed to merge when the galaxies merge within the merger trees, this means that the assumed BH mass depends on the combined total stellar mass of the merged galaxies. Owing to the poor constraints on the BH-bulge relation at high redshift, we rely on the total stellar mass rather than performing a kinematic decomposition of each galaxy's stellar bulge and disk components.

This enables us to make quick BH mass estimates and determine which seeding model realizations produce BH populations in reasonable agreement with empirically derived BHMFs and mass densities. However, it is important to note that this simplified model is by construction most effective at producing a $z = 0$ BH population that matches observations. At high redshift, the BH-galaxy scaling relations are increasingly uncertain, and there are indications that both the scaling relations and their scatter likely evolve with redshift (e.g., Wyithe & Loeb 2003; Booth & Schaye 2011; Pacucci & Loeb 2024). We note also that different empirical measurements of these quantities vary significantly depending on sample selection criteria, especially at high-redshift (e.g., Merloni & Heinz 2008; Shankar et al. 2009; Kormendy & Ho 2013; Reines & Volonteri 2015; Savorgnan et al. 2016). Because the nature of this evolution is not well constrained, in this work we opt for a simplified approach in which a single scaling relation and scatter are used at all redshifts. This simple prescription for mass growth does not attempt to reproduce the wide variety of physical processes that can affect BH accretion rates, including suppression of BH fueling via AGN feedback, BH spin, or dynamical effects (e.g., Chon et al. 2021; Pacucci & Loeb 2022; Inayoshi & Ichikawa 2024). In future work, we plan to undertake a more detailed exploration of BH mass growth prescriptions for our SAM.

2.2.4 Parameter space of BH seeding SAM

We consider a wide selection of seed model realizations that can be divided into two categories based on whether we are systematically varying the minimum threshold for total halo mass ($m_{\rm tot,min}$) or for halo gas mass ($m_{\rm gas,min}$). Each model realization also includes a maximum threshold for the average gas metallicity of halos ($Z_{\rm max}$), and a probability of seeding $f_{\rm seed}$. For each ($m_{\rm tot,min}$, $m_{\rm gas,min}$) pair, we consider six different model realizations with $Z_{\rm max}/Z_\odot = 10^{-1}$, $10^{-1.5}$, or 10^{-2} and $f_{\rm seed} = 0.01$ or 1.

We label these types of seed model realizations as `mgas*_Z*` and `mtot*_Z*` respectively, where the asterisks denote the appropriate values for each parameter. For example, `mgas7_Z0.1` refers to a realization with $m_{\rm gas,min} = 10^7\ M_\odot$ and $Z_{\rm max} = 10^{-1}\ Z_\odot$. Similarly, `mtot8.5_Z0.03` refers to a realization with $m_{\rm tot,min} = 10^{8.5}\ M_\odot$ and $Z_{\rm max} = 10^{-1.5}\ Z_\odot$. In all of our figures, results are presented for each ($m_{\rm gas,min}$, $m_{\rm tot,min}$, $Z_{\rm max}$) combination as a range of values spanning $f_{\rm seed} = 0.01$ - 1. Thus, the value of $f_{\rm seed}$ is not included in the nomenclature. These model realizations and their nomenclature are summarized in Table 1.

3 RESULTS

3.1 Mass-Metallicity Relations of High Redshift Halos

In Figure 1, we examine the distributions of key galaxy properties at high redshift, focusing on gas mass and halo mass versus gas metallicity in the $z = 15$ and $z = 6$ snapshots. We study these galaxy populations to inform the different mass cuts and metallicity cuts that we plan to apply in our SAM-based seed model, as summarized in Section 2.2.4 and Table 1. Low-metallicity galaxies with $Z_{\rm max}/Z_\odot = 10^{-2}, 10^{-1.5}$, or 0.1 make up the majority of hosts at both redshifts. Nearly 100% of halos meet the most lenient metallicity cuts $Z_{\rm max}/Z_\odot = 0.1$ at $z \sim 15$ and $z \sim 6$. Considering the strictest metallicity cuts $Z_{\rm max}/Z_\odot = 10^{-2}$, and discounting halos with no gas at all, the proportions of halos that satisfy this criterion decrease from 97% to 94% from $z = 15$ to 6. The same fraction decreases from 92% to 62% for the star-forming population.

In Figure 3, we investigate the fraction of low-metallicity halos that satisfy the most lenient minimum mass criterion in our seed model: $m_{\rm gas} > 10^7\ M_\odot$ and/or $m_{\rm tot} > 10^8\ M_\odot$. Since nearly all of these galaxies exhibit active star formation, we have excluded an additional star-formation criterion from Figure 3. In the absence of any metallicity criteria (top left panel of Figure 3), the total number of halos meeting these minimum mass criteria grows from a few $\times 10^4$ at $z = 15$ to $\gtrsim 10^6$ by $z = 6$. Star formation, feedback processes, and

Type of Model Realization	Name of Model Realization	$m_{\rm tot,min}$ [$\log_{10}\ M_\odot$]	$m_{\rm gas,min}$ [$\log_{10}\ M_\odot$]	$Z_{\rm max}$ [$\log_{10}\ Z_\odot$]	$f_{\rm seed}$ -
Varying $m_{\rm gas,min}$	`mgas[7,8,9]_Z[0.01,0.03,0.1]`	8.0	(7.0, 8.0, 9.0)	(-2.0, -1.5, -1.0)	(0.01, 1.0)
Varying $m_{\rm tot,min}$	`mtot[8.5,9.5,10.5]_Z[0.01,0.03,0.1]`	(8.5, 9.5, 10.5)	7.0	(-2.0, -1.5, -1.0)	(0.01, 1.0)

Table 1. Summary of semi-analytic BH seeding model realizations used in this work. For each classified type ("varying $m_{\rm gas,min}$" or "varying $m_{\rm tot,min}$"), we consider three values of the relevant mass threshold (while keeping the other mass threshold fixed), as well as three values of $Z_{\rm max}$ and two values of $f_{\rm seed}$. Our SAM therefore includes 36 distinct BH seeding realizations. Names of model realizations specify the variable mass threshold and the metallicity threshold: `mgas*_Z*` or `mtot*_Z*`. $f_{\rm seed}$ is not included in the nomenclature, as all results are presented as a range of values when $f_{\rm seed}$ is varied from 0.01 to 1.

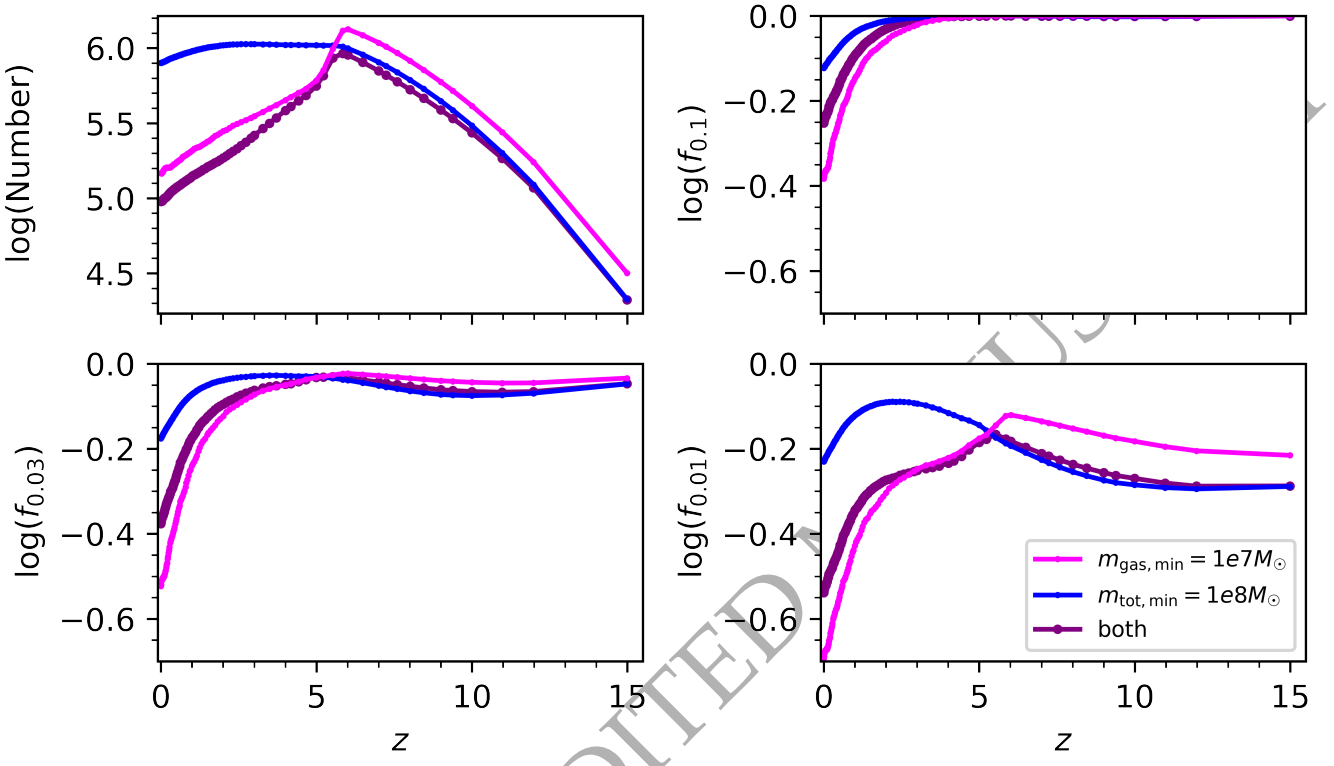

Figure 3. The evolution of metal-poor halo sub-populations is shown. In the top left panel, we show the total number of halos above a minimum mass of $m_{\rm tot,min} = 10^8\ M_\odot$ (in blue), $m_{\rm gas,min} = 10^7\ M_\odot$ (in magenta), or both (in purple). All other panels show the fraction of these halos that are metal-poor as defined via their average gas-phase metallicity, using the same color scheme. The top right panel shows the fraction of halos that satisfy $Z_{\rm max}/Z_\odot = 0.1$, which we denote as $f_{0.1}$, and the bottom left and bottom right panels show the fractions with $Z_{\rm max}/Z_\odot = 10^{-1.5}$ or 10^{-2}, denoted as $f_{0.03}$ and $f_{0.01}$, respectively. We see that $f_{0.1}$ remains very high until $z \sim 4$ and then declines as halo enrichment proceeds towards cosmic noon. Stricter metallicity cuts show similar trends but are also modulated by the increase in total number of halos up to $z \sim 6$.

mergers subsequently reduce the number of halos meeting the gas mass criterion after $z \sim 6$. By $z = 0$, there are roughly 8×10^5 halos that satisfy $m_{\rm tot} > 10^8\ M_\odot$, 1.5×10^5 halos that satisfy $m_{\rm gas} > 10^7\ M_\odot$, and 9.5×10^4 halos that satisfy both criteria.

The remaining three panels in Figure 3 show the fraction of these halos that meet not only the specified mass cuts but also satisfy the maximum metallicity cuts $Z_{\rm max}/Z_\odot = 10^{-2}, 10^{-1.5}$, or 0.1, denoted as $f_{0.01}, f_{0.03}$, and $f_{0.1}$, respectively. The top right panel shows that nearly all of these halos have $Z_{\rm max}/Z_\odot = 0.1$, through the epoch of reionization. It is only at redshifts below $z \sim 4$ that the metal-poor fraction noticeably declines, as the Universe approaches the peak of star-forming activity at "cosmic noon." For the population that satisfies both mass cuts, $f_{0.1}$ goes from nearly 100% at $z \sim 15$ to 56% at $z \sim 0$.

With a stricter metallicity cut of $Z_{\rm max}/Z_\odot = 10^{-1.5}$, we see broadly similar behavior with some minor differences. Roughly 90% of these halos are below this enrichment level at $z \sim 15$. We also see a slight temporary dip in the fraction of metal-poor halos between $z = 15$ and $z \sim 6$, owing to the interplay between halo enrichment via star formation and the steady increase in the total number of metal-poor halos meeting the mass cuts. Below $z \sim 6$, the number of halos levels out

and eventually declines due to mergers, while the number of halos above $m_{\rm gas,min}$ sharply declines owing to a burst of star formation and feedback. After this point, continued metal enrichment steadily decreases the fraction of metal-poor halos. These trends are starker for the lowest metallicity threshold $Z_{\rm max}/Z_\odot = 10^{-2}$. Only about 52% of halos exceeding both mass thresholds lie below $Z_{\rm max}/Z_\odot = 10^{-2}$ at $z = 15$, and by $z = 0$ this metal-poor fraction is 29%.

3.2 SAM verification: Reproducing the TNG BH population

Before we attempt to explore the physically motivated, SAM-based seed model with realizations summarized in Section 2.2.4, we first verify that our approach can successfully reproduce the actual TNG results when the TNG seeding criterion is applied. We impose a minimum halo mass of 5×10^{10} $M_\odot h^{-1}$, consistent with the TNG BH seeding criterion. Figure 4 compares our TNG-analogue semi-analytic seeding prescription to the true number density of BHs within the CSHs in the TNG simulation. Because of the CSH proxies used in the model, number density evolution is compared with that from TNG CSH BHs, but there is little difference between the halo and CSH model results; this means that the population of satellite galaxies in TNG that meet the model seeding criteria and host BHs is small (see Appendix A1). The model agrees well with the actual number density of BHs in TNG at all redshifts; at $z = 0$, the model agrees with TNG to within 4%.

In Figure 4, model results are also compared with empirical BH number densities (left panel) and mass densities (right panel) (Merloni & Heinz 2008; Shankar et al. 2009; Cao 2010; Shen et al. 2020, hereafter referred to as M08, S09, C10, and S20, respectively). M08 and C10 both use the BHMF continuity equation but make different assumptions about the growth of the BHs. M08 empirically determine the Eddington ratio distribution by coupling the empirical BH mass function and X-ray luminosity function with fundamental relations between three different accretion mode observables, while C10 assumes a power-law Eddington ratio distribution. S09 models AGN and SMBH populations under the assumption that the BHMF grows at the rate implied by the observed luminosity function. S20 give updated constraints on the bolometric quasar luminosity function from observations from the past decade with an updated quasar SED model and bolometric and extinction corrections. At $z \sim 0$, these studies predict BH number densities ranging from $1.3 - 4.2\times10^{-2}$ cMpc^{-3}; the TNG $z = 0$ number density of $n_{\rm BH} = 1.96\times10^{-2}$ lies in the middle of these values. These studies predict mass densities with much less variation at $z \sim 0$, ranging from $4.13 - 4.88\times10^5$ $M_\odot$ cMpc^{-3}, whereas the local TNG mass density of $\rho_{\rm BH} = 8.88\times10^5$ $M_\odot$ cMpc^{-3} lies above those values. In our SAM realization matched to TNG seeding criteria, the mass densities agree well at $z \lesssim 2$, but they are somewhat overpredicted with respect to actual TNG before cosmic noon, likely due to non-evolution of the scaling relation in the model.

Notably, there are substantial discrepancies between the different empirical constraints on the BH number density, which increase at higher redshift. There are also significant discrepancies between the TNG and the empirically estimated BH number densities, especially at high redshift. Previous studies have similarly found that although the low-redshift TNG QLFs and the $z = 0$ BHMFs agree reasonably well with observations (Sijacki et al. 2015; Weinberger et al. 2018), TNG overpredicts the high-redshift QLF (Weinberger et al. 2018). Other simulations using similar physical models have also been found to overpredict the bright end of the AGN luminosity function at high redshift (Bhowmick et al. 2021). However, high-redshift quasar statistics remain incomplete and poorly constrained, particularly at the faint end of the luminosity function. This creates large uncertainties in the BHMF at early times, especially at the low mass end.

JWST has already uncovered substantial new populations of AGN at high redshifts (e.g., Onoue et al. 2023; Larson et al. 2023; Maiolino et al. 2023; Kocevski et al. 2023; Maiolino et al. 2024) and will transform our understanding of the high-redshift AGN luminosity function in the coming years. Advances in theoretical models of high-redshift BH populations will be crucial for interpreting this new wealth of data from JWST, and in preparation for LISA observations of the high-redshift GW Universe. The large BH seed masses used in many simulations ($\sim 10^6 M_\odot$ in TNG) likely contribute to overestimation of the low mass end of the BHMF at high redshift, but at the same time, observational constraints on low-mass, high-redshift BHs are highly incomplete. This is precisely one of the issues that our present work addresses by modeling BH populations with lower seed masses and a much wider range of seeding criteria.

The host mass histograms in the left panels of Figure 5 at $z = 0$ and $z = 3$ show that not only does the total number of BHs agree well between our SAM and TNG, but also the distribution of host halo masses. Note that in both cases, a tail of BH host masses extends below the minimum required halo mass for BH seeding in TNG (5×10^{10} $M_\odot h^{-1}$), especially at $z = 0$. These are galaxies that have lost mass over time via tidal stripping.

3.3 Fiducial Semi-Analytic BH Seeding Criteria

Having validated our SAM by reproducing the TNG results, we are now finally ready to explore the seeding criteria outlined in Section 2.2.4. In Figures 6, 7 and 8, we analyze BH populations produced by the seed model in terms of their number density and mass density evolution. We consider two distinct types of BH populations:

- The *full population* of BHs formed in our SAM realizations, referred to as the "FP BHs". With all of the masses determined via the local BH scaling relations, the FP BHs have masses ranging from $\sim 10^3$ $M_\odot$ to $\sim 10^{10}$ $M_\odot$. The lower BH mass limit is set by the adopted BH-stellar mass scaling relation and the requirement that the stellar mass be nonzero.
- BHs with masses $> 10^5$ $M_\odot$, hereafter referred to as the *massive population* of BHs or "MP BHs."

In the following subsections, we will systematically address the impact of our seed model on different aspects of the number density and mass density evolution of the resulting BH populations.

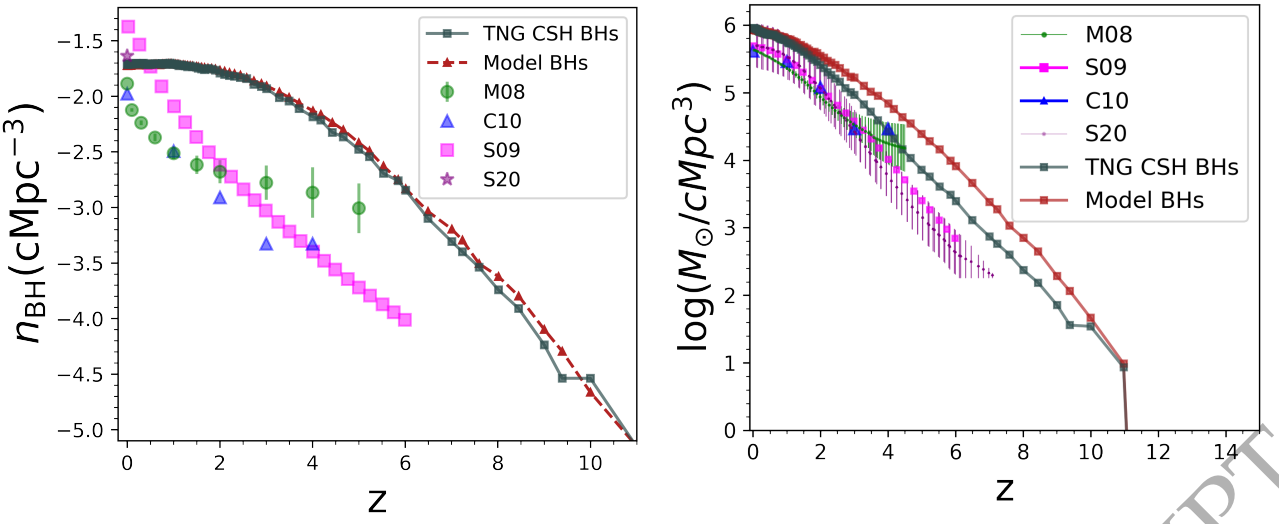

Figure 4. We apply the TNG halo criterion 5×10^{10} $M_\odot h^{-1}$and choose a CSH proxy for these eligible halos for our halo model realizations, assuming one BH per halo. We compare the halo model results to TNG CSH BHs due to our model choice of CSH proxies. In the left panel, the number density ($n_{\rm BH}$) of BHs (in units of comoving Mpc^{-3}) in our TNG model (in the dashed red line) comes close to that of TNG CSHs (in the solid dark gray line) at all redshifts and to within 4% at $z = 0$. $n_{\rm BH}$ from the empirical studies M08 (green circles), C10 (blue triangles), S09 (magenta squares), and S20 (purple star) are shown (see full references in § 3.2). At $z \gtrsim 0.5$, TNG predicts higher number densities than observations, but it lies squarely in the middle of the empirical data at $z = 0$. In a similar fashion, the right panel shows the evolving total BH mass density of the TNG model with the TNG BH mass-total stellar mass relation as the growth model. The model mass density at $z = 0$ agrees with TNG to within 2%. The SAM and TNG BH mass densities agree well with each other at $z \lesssim 2$, but predict higher mass densities than empirical constraints at all redshifts.

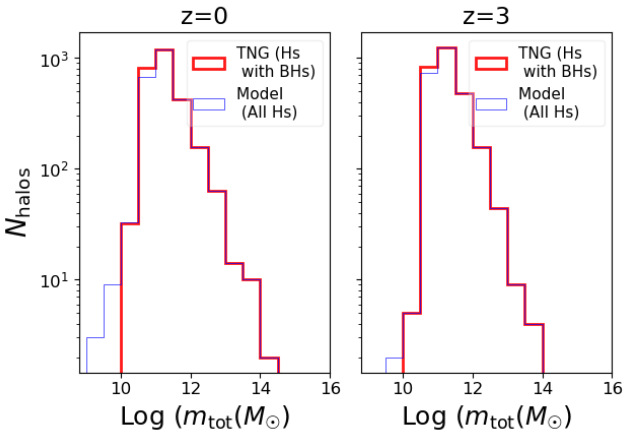

Figure 5. The total mass distributions at $z = 0$ and $z = 3$ are shown for halos that have BHs in TNG (red) versus the halos hosting BHs in our TNG-analogue model realization (blue). Both histograms show close agreement between host masses from the model and TNG halos.

3.3.1 Redshift evolution of BH number densities

The shaded regions in Figure 6 shows the number densities of FP (cool colors) and MP (warm colors) BHs predicted by our seed model. Different colors indicate model realizations with different $Z_{\rm max}$ values, and the shaded region spans stochastic seeding values from $f_{\rm seed} = 0.01$ to 1. As expected, number densities increase quickly with time at the highest redshifts, as rapid halo growth drives the formation of new seeds. For most model realizations, the number densities peak at redshifts between $\sim 2-7$, when halo enrichment slows the formation of new seeds, after which it decreases with time. This is due to a combination of several effects as identified in Section 3.1: 1) seed formation is slowed by metal enrichment in halos, 2) star formation and feedback can reduce the amount of gas available to form seeds inside the halos, 3) the BHs undergo mergers with each other.

Generally, we see that the saturation in the BH number densities tends to happen at later times as the seeding criteria become more strict. This happens because of a combination of effects. First, the maximum number of halos (with masses $> 10^7$ $M_\odot$) available for seeding saturates at $z \sim 6$ (revisit Figure 3, blue line). This essentially sets $z \sim 6$ to be the "saturation redshift" of the FP BH number densities for the most lenient seed model realizations like `mgas7_Z0.1` and `mtot8.5_Z0.1` (Figure 6, leftmost panels). But in stricter seed model realizations, the BH occupation fraction in halos is lower at early times ($z \gtrsim 6$), such that proportionally more halos are available to form new seeds at $z \lesssim 6$. Therefore, for the stricter seeding criteria like `mgas9_Z0.1` and `mtot10.5_Z0.1` (Figure 6, rightmost panels), the saturation in the FP BH number densities starts to occur at lower redshifts (i.e., $z \sim 3$ and $z \sim 2$, respectively). Irrespective of these trends, however, we see that more lenient seed model realizations form more BHs at all redshifts.

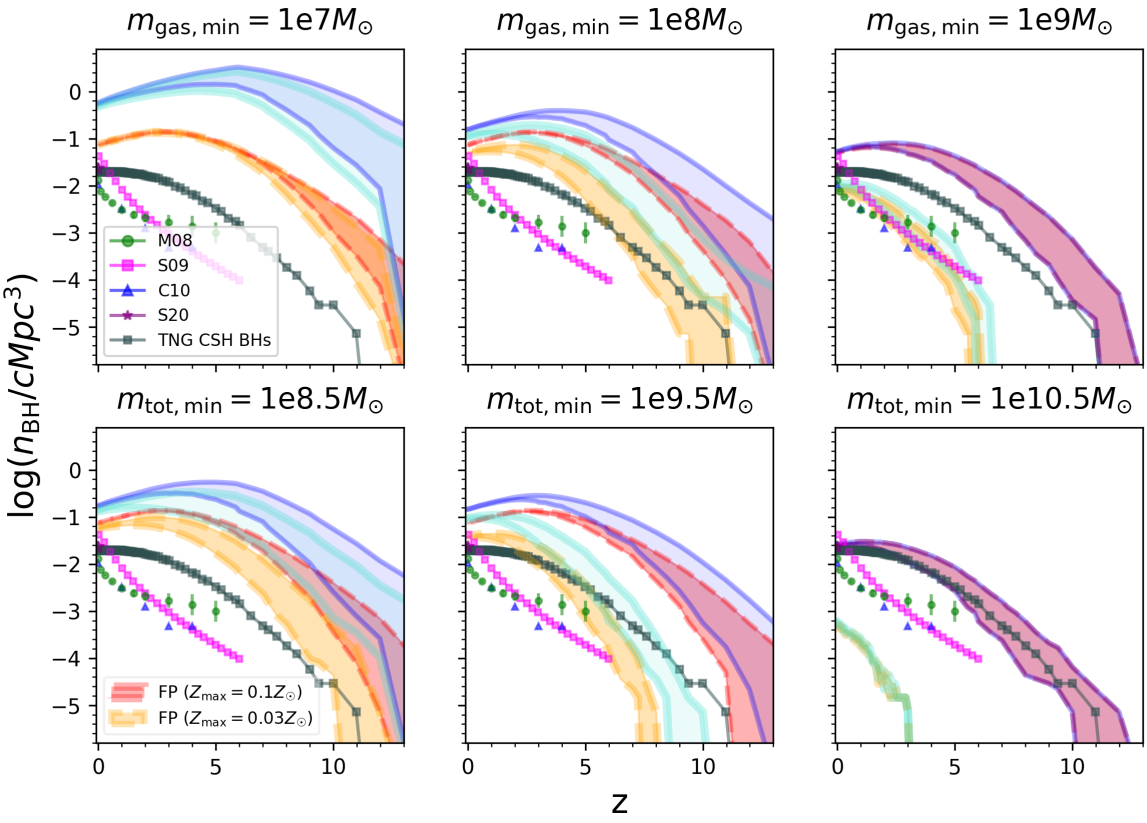

Figure 6. Comoving BH number densities, $n_{\rm BH}$, are shown versus redshift for fiducial seeding criteria. FP and MP results are shown in cool and warm-colored transparent shaded regions, respectively. The lower and upper limits of the shaded regions correspond to probabilistic seeding fractions between 0.01 and 1. The model realizations each have different gas mass, host mass, and metallicity criteria ($Z_{\rm max} = 0.1 Z_\odot$ and $10^{-2} Z_\odot$) (in red and gold, respectively, for the MP, and in blue and turquoise, respectively, for the FP). The top panels correspond to varying-$m_{\rm gas,min}$ realizations and the bottom panels correspond to varying $m_{\rm tot,min}$. The parameters were chosen systematically and not with the intent of producing the closest fit. Several different SAM realizations span number densities that agree with results from TNG and AGN observations (M08, C10, S09, and S20). At high redshifts, BHs $< 10^5$ $M_\odot$ are the most significant contributors to the number densities by factors of $\sim 10 - 100$. This underscores the importance of LISA's capabilities to detect these systems at high redshift.

3.3.2 Impact of halo mass and gas mass seeding thresholds on the BH number densities

Not surprisingly, the BH number densities tend to decrease with increasing halo mass ($m_{\rm tot,min}$) and gas mass seeding thresholds ($m_{\rm gas,min}$). The impact is generally stronger at higher redshifts, simply because the underlying halo mass functions are steeper. Additionally, the FP BH number densities seem almost equally sensitive to the seeding criteria as the MP BH number densities. For example, as we go from the most lenient to the strictest $m_{\rm gas,min}$, (i.e., $m_{\rm gas,min} = 10^7$ to 10^9 $M_\odot$), both the FP and MP BH number densities can be suppressed by one to two orders of magnitude for $Z_{\rm max} = 0.1$ (blue versus red shaded regions in Figure 6, top panels). However, for the majority of the halo and gas mass thresholds, the BH population is dominated by the FP BHs. These are largely comprised of low-mass ($\sim 10 - 10^5$ $M_\odot$) BHs that are currently inaccessible to EM observations at high redshift. However, upcoming GW facilities like LISA will be sensitive to low-mass, high-redshift mergers, which will likely provide strong constraints on seed models.

3.3.3 Impact of gas metallicity threshold on the BH number densities

We now compare the number density predictions of MP and FP BHs for two different gas metallicity thresholds for seeding, i.e., $Z_{\rm max} = 0.1$ & 0.01. We can clearly see that when the halo and gas mass thresholds are increased, the metallicity threshold has a stronger impact on seeding. For example, among the varying-$m_{\rm gas,min}$ model realizations (blue vs. turquoise regions in the top panels of Figure 6), when $m_{\rm gas,min}$ is 10^7 $M_\odot$, decreasing $Z_{\rm max}$ from 0.1 to 0.01 $Z_\odot$ makes a very small difference in the number densities of FP BHs. For a higher $m_{\rm gas,min}$ of 10^9 $M_\odot$, $Z_{\rm max} = 0.01$ $Z_\odot$ produces up to ~ 1000 times fewer BHs compared to $Z_{\rm max} = 0.1$ $Z_\odot$. Overall, this is because more massive halos tend to be more metal enriched due to a more extensive history of star formation and evolution. As we can see in Figure 1, the vast majority of $> 10^7$ $M_\odot$ halos have metallicities < 0.01 $Z_\odot$. In contrast, a very small minority of $> 10^{10}$ $M_\odot$ halos have metallicities < 0.01 $Z_\odot$.

The impact of the metallicity criterion substantially de-

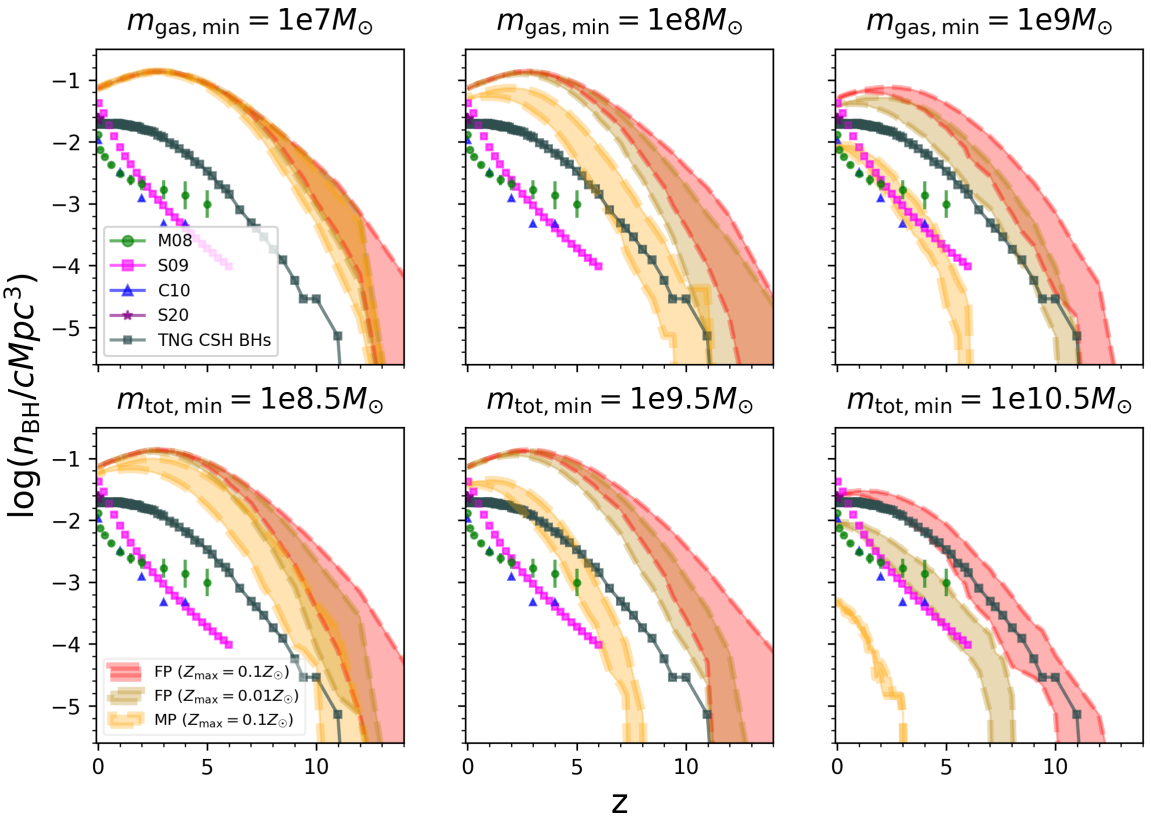

Figure 7. Comoving BH number density evolution is shown in the same format as Figure 6, except only the MP model results are shown (i.e., all results include only BHs $> 10^5\ M_\odot$). In addition, SAM realizations with intermediate host metallicity thresholds of $Z_{\rm max}/Z_\odot = 10^{-1.5}$ are shown in the tan shaded region. Numerous model realizations fall within the empirical constraints and their considerable observational uncertainties by $z = 0$ with most of the model realizations overpredicting with respect to TNG before then. Recall, however, that the MP results include all BHs above $10^5 M_\odot$, whereas TNG BHs have a minimum mass of $8 \times 10^5 M_\odot$. The fact that the model spans the empirical space illustrates its capabilities to explore more realistic seed mass variations by seeding in lower-mass hosts than the TNG halo seed mass threshold.

creases with time in general. In fact, for the lowest $m_{\rm tot,min}$ and $m_{\rm gas,min}$, model realization criteria with different $Z_{\rm max}$ values produce similar results at $z \sim 0$. This trend is not surprising because even though more BHs are formed at earlier times in model realizations with higher $Z_{\rm max}$, cosmic evolution causes them to merge with each other as their host halos merge. As a result, the differences in the high-z number densities seen for model realizations with different $Z_{\rm max}$, washes out over time. To that end, note that our model assumes prompt mergers amongst BHs within the same halo, not excluding wandering off-center BHs or BHs in satellite galaxies that any of the central subhalo proxies evolve into. We could expect the impact of $Z_{\rm max}$ to behave differently given a stronger variety of host subhalos (see Appendix A). But we reiterate the model's flexible nature in which the assumption of one BH per halo leaves room for seeding to not be in mostly central subhalos as long as the descendants in the TNG merger trees are followed in the model.

3.3.4 Impact of seed probability on the number densities

Here we examine the impact of probabilistic seeding ($f_{\rm seed}$) on the BH number densities. Note that the seed probability is applied (as a random draw) on every descendant along a given tree. In the absence of a metallicity criterion for seeding, applying such a probabilistic seed criterion would simply lead to an effective delay in the seed formation along a tree branch. However, the presence of a metallicity criterion dictates that the formation of a seed on a tree branch hinges upon the rate of metal enrichment along that branch. If the tree branch undergoes rapid metal enrichment and the seed probability is low enough such that the branch is already enriched with metals by the time sufficient random draws are available to place a seed, then no seed will form on that particular branch at all. BH number density predictions for seed probabilities of 1 and 0.01 are shown as the upper and lower limits of the shaded regions in Figure 6. We can see that the shaded regions tend to shrink as redshift decreases; in fact, by $z \sim 0$, both seed probabilities produce very similar number densities. This suggests that for a seed probability of 0.01, metal enrichment does not occur rapidly enough to completely prevent seeding on the vast majority of the tree

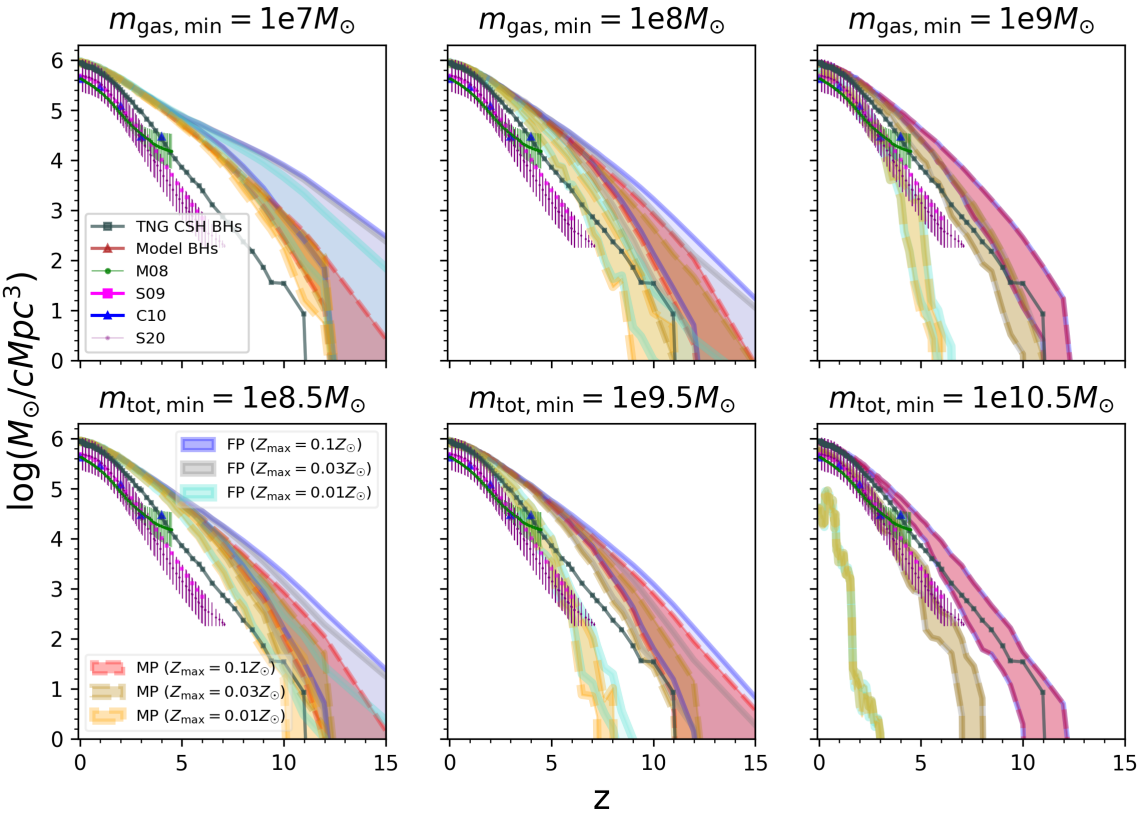

Figure 8. Comoving mass densities are shown for seeding model realizations in the same style and corresponding to those in Figure 7. The model realizations in closest agreement with empirical constraints at $z \lesssim 6$ generally do not start forming seeds until $z < 10$. Model realizations that do produce high-z BHs tend to predict higher mass densities at intermediate redshifts than TNG or empirical data, in part because our predictions include $< 10^6 M_\odot$ BHs. Among these, `mgas8_Z0.01` and `mgas9_Z0.03` show the best agreement with TNG and empirical predictions. Differently from the number density results, the empirical constraints on mass density are in much closer agreement with each other and with our seed model realizations. `mtot10.5_Z0.01` is the only realization that severely underpredicts the mass densities. It does not start producing seeds until $z \sim 3$.

branches. We emphasize that the seeding probability is applied at every snapshot (every step in the merger tree) to all galaxies meeting the other seeding criteria, with no minimum redshift (cf. the probabilistic seeding models of Buchner et al. 2019; Spinoso et al. 2023). As a result, the cumulative probability of a galaxy being seeded by $z = 0$ is much higher than the value of $f_{\rm seed}$. Mergers between galaxies also tend to increase the BH occupation fraction over time, though mergers will be most important for massive halos.

It is useful to compare the impact of seed probability vs that of gas metallicity threshold since the former is intended to account for additional physics (halo growth, star formation and metal enrichment) that can influence seeding. Notably, we find that the impact of reducing $Z_{\rm max}$ from 0.1 and 0.01 on the number densities is stronger than that of reducing $f_{\rm seed}$ from 1 to 0.01. Nevertheless, both parameters do have a significant impact, particularly at the highest redshifts. This motivates the need for exploring the variety of other physics that can impact BH seeding such as UV radiation, gas angular momentum, dynamical heating, etc.; this will be focus of future work. Note that because $f_{\rm seed}$ is the seeding probability applied at each snapshot, it implicitly depends on the time resolution of TNG snapshots and those spanned by the galaxy merger trees. In other words, a given value of $f_{\rm seed}$ would not have the same physical meaning in a simulation with higher or lower time resolution.

We emphasize that for a given choice of $f_{\rm seed}$, the limited snapshot resolution of TNG does not have a significant impact on our results. It is possible that a minor suppression of seed formation occurs in halos that become metal-enriched above $Z_{\rm max}$ between snapshots but did not meet the minimum mass criterion in the previous snapshot. Aside from this, however, the BHs in our model are evolved as passive tracers of the underlying galaxy merger tree and stellar mass. Thus, if a model realization formed a seed at $z = 15$ that with higher snapshot frequency might have formed at $z = 18$, for example, our model would still report the same BH mass distribution and occupation fraction at $z = 15$.

3.3.5 *Comparison of the number density predictions to TNG and empirical data*

We finally compare the BH number densities predicted by our seed model to empirical data shown in Figure 6. Note that

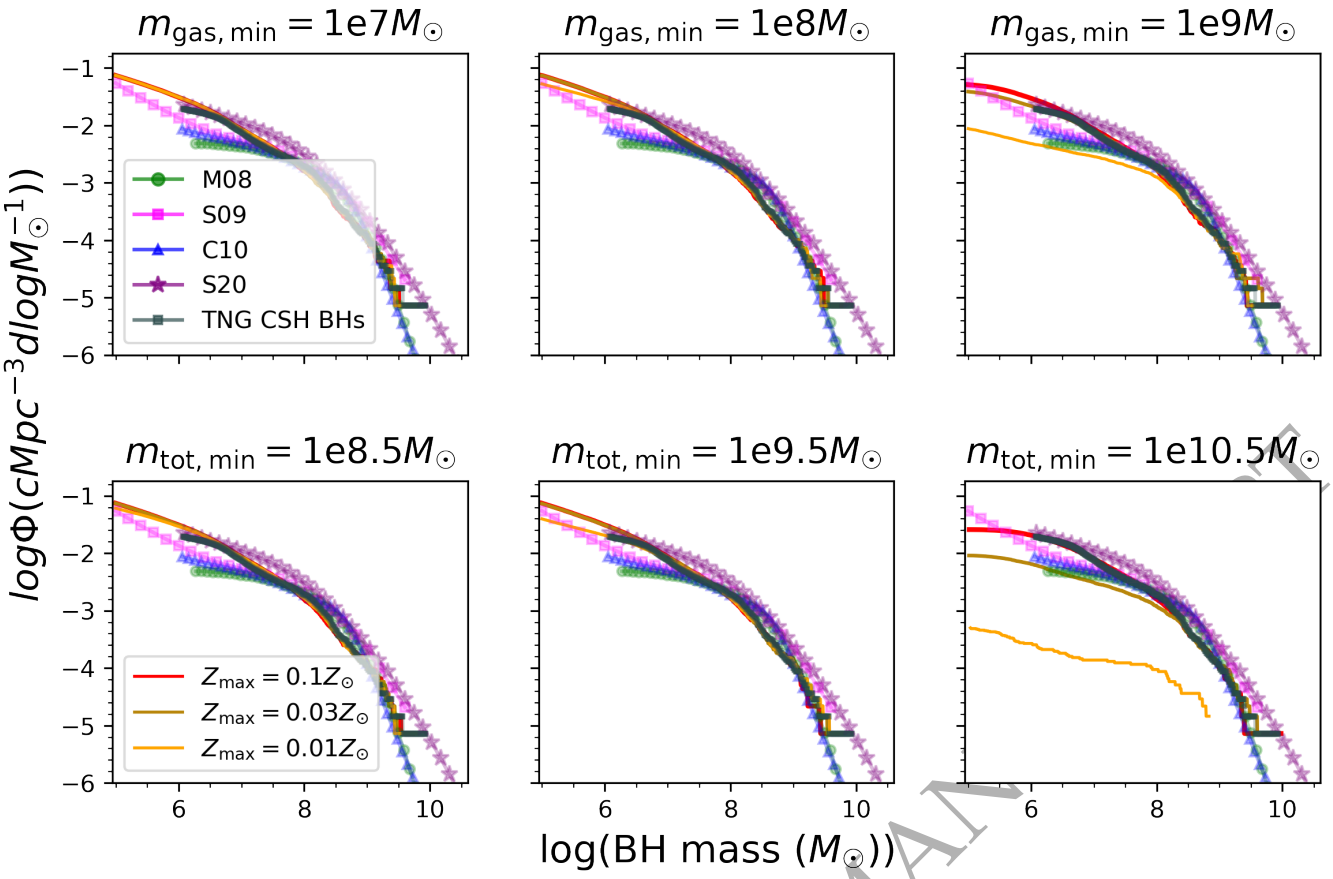

Figure 9. In the same color scheme as Figure 8, the corresponding $z = 0$ BHMFs are shown for the model. There are well-produced local BHMFs for all the plausible SAM realizations that are well within the empirical parameter space and even those that are not. Above ($\gtrsim 10^6\ M_\odot$), many model realization BHMFs are nearly identical to the TNG BHMF. At the most massive end ($\gtrsim 10^9\ M_\odot$), nearly all model realizations are in best agreement with M08 until it produces a lower "knee" and with the exception of ones that produce too few BHs overall. $Z_{\rm max}/Z_\odot < 10^{-2}$ hosts at $z = 0$ are more rare and do not contribute to the most extreme mass end of the BHMF in the lower right panel. BHMF results are as expected with the BH - stellar mass scaling relation model with scatter and with the well-produced stellar mass function of TNG.

observations have thus far not been able to probe BH populations $\lesssim 10^5\ M_\odot$. Therefore, it is not surprising that for most of our seed model realizations, the number densities of FP BHs substantially exceed that of the empirical data. To that end, the MP BHs offer a fairer comparison to the empirical data. However, recall from Section 3.2 that even amongst the empirical data, the various published measurements vary by factors of up to ~ 4 at $z \sim 0$. Given the uncertainties in both our model and in empirical data, the following serves merely as a broad comparison between simulations and observations.

In Figure 7, we replot the predicted BH number densities already shown in Figure 6, but here we solely focus on the MP BHs. We also include an additional, intermediate model realization with $Z_{\rm max} = 0.03\ Z_\odot$. The most lenient criteria as in `mgas7_Z0.1` and `mtot8.5_Z0.1` predict BH number densities that differ from empirical constraints at $z \sim 0$ by factors of up to ~ 6. Note that only S09 attempt to include $10^5 - 10^6\ M_\odot$ BHs in their analysis; the others use $10^6\ M_\odot$ as the lower limit on BH mass. At higher redshifts, the empirical constraints become even more uncertain. This undoubtedly contributes to the fact that a substantial majority of our model realizations predict higher BH number densities than are obtained from empirical constraints, with the exception of S09 at $z \sim 0$. This includes the `mgas9_Z0.01` and `mtot10.5_Z0.1` model realization for which the number density evolution most closely resembles that from TNG.

In fact, only two of our (strictest) model realizations predict BH number densities within the range spanned by the different empirical constraints at intermediate redshifts of $z \sim 1 - 5$; these are `mgas9_Z0.01` and `mtot10.5_Z0.03` (top right and bottom right panels, respectively, in Figure 7). However, `mtot10.5_Z0.03` underpredicts the $z = 0$ number densities, and both model realizations do not begin forming BHs until $z < 8$, which is inconsistent with recent discoveries of very high-redshift AGN with JWST (Larson et al. 2023; Onoue et al. 2023; Scholtz et al. 2023; Fujimoto et al. 2022; Maiolino et al. 2024). This underscores the importance of low-mass BHs in calculations of the BH number density. This is even more true as we go to higher redshifts, where the empirical constraints become increasingly uncertain. Nevertheless, we can still fully rule out one of our strictest model realizations (i.e., `mtot10.5_Z0.01`), which predicts number densities substantially below all of the empirical constraints

at all redshifts; this is because the seed production does not start until $z \sim 3$ (bottom right panels of Figures 6 and 7).

Overall, the foregoing results demonstrate that BH number densities are sensitive to different seeding scenarios, particularly at higher redshifts wherein the variations within our seed model are large (exceeding ~ 100 and for both FP and MP BHs. This is consistent with previous studies that find orders of magnitude variation in the predicted number density of massive BH seeds, particularly at high redshift (e.g., Agarwal et al. 2012; Dijkstra et al. 2014). These works also find that imposing restrictive DCBH seeding conditions can dramatically reduce seed formation.

As JWST continues to uncover surprising new populations of high-redshift AGN, the parameter space of possible BH formation channels will become more constrained. Additionally, observations of even lower mass ($\sim 10^4 - 10^6\ M_\odot$) BHs with upcoming LISA and proposed EM facilities such as Lynx can pose even more stringent constraints on our seed model. At the other end, our predicted number densities may also be impacted by the modeling of physical processes such as star formation, metal enrichment and stellar feedback. In future work, we will continue to use our newly built framework to systematically explore the impact of all these processes. This will be crucial preparation for the wealth of observational data that we expect from the coming decades.

3.3.6 Mass density evolution

Mass density evolution (Figure 8) varies much less between the seed model realizations, compared to their number density evolution. This is especially true at $z \lesssim 4$, where both MP and FP BHs converge to similar values. Even in seed model realizations with the most lenient total and gas mass cuts, for which number densities are dominated by BHs $< 10^5\ M_\odot$, we still see similar $z = 0$ mass densities between the MP and FP BHs. This implies that for all of our model realizations, the mass densities at $z \lesssim 4$ are dominated by BHs significantly more massive than $10^5\ M_\odot$. Notably, the empirical constraints on mass density (most of which extend to $z \lesssim 4$) are also in much closer agreement with each other compared to the number densities.

At $z \gtrsim 1$, the mass densities from empirical constraints, from TNG, and from our model begin to diverge from one another. The different model realizations tend to fall into one of two categories: either they are so restrictive that they do not begin forming seeds until $z < 10$, or they predict higher mass densities at high redshift, relative to empirical constraints and TNG. While this may in part reflect a true overprediction of BH mass densities, recall that our MP is a complete sample of BHs down to $10^5 M_\odot$—a mass regime where observational samples are highly incomplete owing to selection biases and the intrinsic faintness of low-mass AGN. Empirical constraints are also increasingly uncertain beyond the local universe, and the most recent observational evidence from JWST strongly indicate that high-redshift BHs are overmassive relative to their hosts (Pacucci et al. 2024). We emphasize that our model results are also subject to uncertainty and biases produced by our assumptions, which must be considered as well when comparing to empirical data.

In terms of which seeding parameter combinations produce the closest agreement with TNG and with empirical constraints, without being so restrictive as to start seed formation at late times, `mgas8_Z0.01` and `mgas9_Z0.03` are arguably the two most plausible model realizations. For this reason, we focus on the $m_{gas,min} = 10^8 M_\odot$ and $m_{gas,min} = 10^9 M_\odot$ model variations for more detailed examination in subsequent analysis.

The only seed model realization that severely underpredicts the mass densities is `mtot10.5_Z0.01`, since it does not start producing seeds until $z \sim 3$. Recall here that the BH masses are based on the local $M_* - M_{\rm BH}$ TNG scaling relation. Therefore the agreement between the seed model realizations and the empirical measurements is not surprising, given that the underlying TNG galaxy formation model successfully reproduces the observed stellar mass functions and the cosmic star formation rate densities at $z \lesssim 4$ (Genel et al. 2018). In this sense, the agreement between model and empirical mass densities at $z = 0$ can be viewed as a validation, rather than a prediction, of our model.

At $z \gtrsim 4$, we see more variations in the mass density predictions between the different seed model realizations. In this regime, the only available empirical constraints are from S20 that span from $z \sim 0 - 7$. Future constraints on the mass densities at high redshifts using facilities like JWST will help us better discriminate between the different seeding scenarios. Finally, we also note that at $z \gtrsim 4$, the mass densities for the FP BHs are significantly higher than the MP BHs. This implies that at these redshifts, the overall mass densities are largely contributed by low-mass ($< 10^5\ M_\odot$) BHs, which will be difficult to access with EM observatories. LISA observations are therefore going to play an essential role in constraining the mass densities at these redshifts.

3.4 Local BHMFs Produced by the Fiducial BH Seeding Criteria

The $z = 0$ BHMFs for our model realizations are shown in Figure 9 with the same color scheme as Figure 8. The probabilistic seed model realizations are omitted since they produce the same results as the non-probabilistic model realizations by $z = 0$. There is very little variation in the predicted BHMFs among most of our model realizations. Overall, our seed model BHMFs are in broad agreement with the empirical BHMFs. As we make the seeding criteria more restrictive, we see underpredictions of the BHMF values, starting at the lowest mass end (right panel). For the strictest seeding criteria (lower right panel), we see appreciable variations in the BHMF over a larger range of BH masses. Not surprisingly, there is generally more variation at the low mass end due to greater retention of the memory of the initial seed mass. Our seed model predicts similar BHMFs as the TNG model even at the low mass end close to the TNG seed mass ($\gtrsim 10^6\ M_\odot$). At the most massive end ($\gtrsim 10^9\ M_\odot$), both TNG and nearly all of our seed model realizations lie within the range of the empirical BHMFs; this is with the obvious exception of `mtot10.5_Z0.01` that produces too few seeds for any significant BH population to form. Below $\sim 10^6\ M_\odot$, `mgas9_Z0.01` underpredicts with respect to the BHMF values of S09. All of our seeding model realizations do, however, produce a nearly identical "knee" in the BHMF compared with that of TNG, but slightly lower compared to within the observational BHMFs.

In any case, it is fair to say that given the spread within the empirical BHMFs themselves, most of our seed model re-

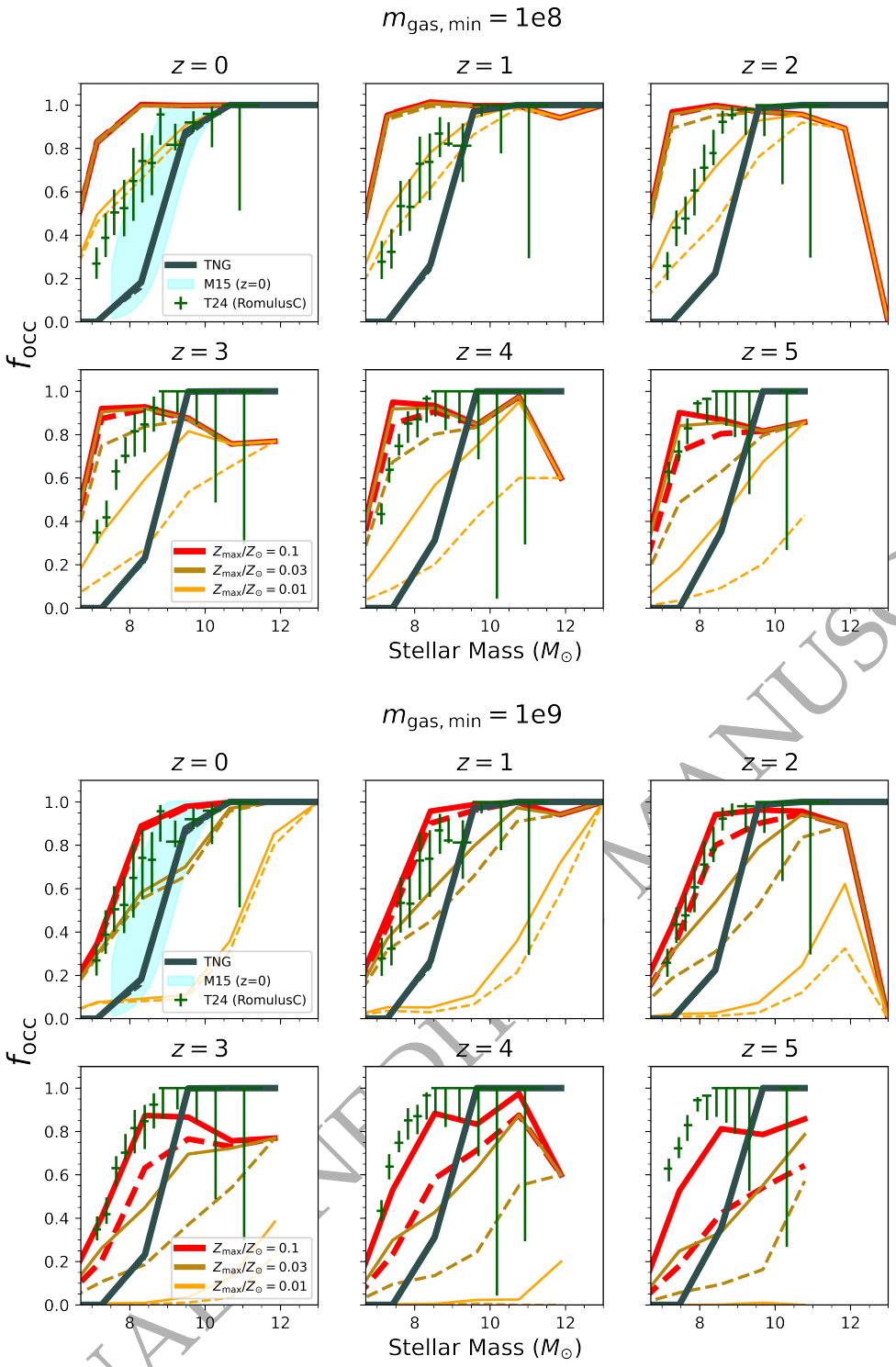

Figure 10. The $z = 0 - 5$ BH occupation fractions are shown for all metallicity and $f_{\rm seed}$ variations (in the same previous colors and linestyles) and for MP model realizations with $m_{\rm gas,min} = 10^8 M_\odot$ (top panel) and $m_{\rm gas,min} = 10^9 M_\odot$ (bottom panel) compared with TNG results (grey-green solid line), the $z = 0$ $f_{\rm occ}$ of the optically selected early-type galaxy clean sample from M15 (within 1σ confidence; light blue shaded region), and the $f_{\rm occ}$ from dwarf galaxies in the RomulusC simulation (T24). At $z = 0$, the BH occupation fractions in the most massive galaxies are unity for all model realizations, and for TNG. At all redshift, most of our model realizations have higher occupation fractions than TNG for low-mass galaxies, reflecting the more lenient seeding criteria in our model realizations and producing better agreement with M15 and T24. Note that $f_{\rm occ} < 1$ in the highest-mass bin at $z = 2 - 4$ owing to a single massive halo in TNG50 that is metal-enriched very early and has a quiet merger history until $z < 2$.

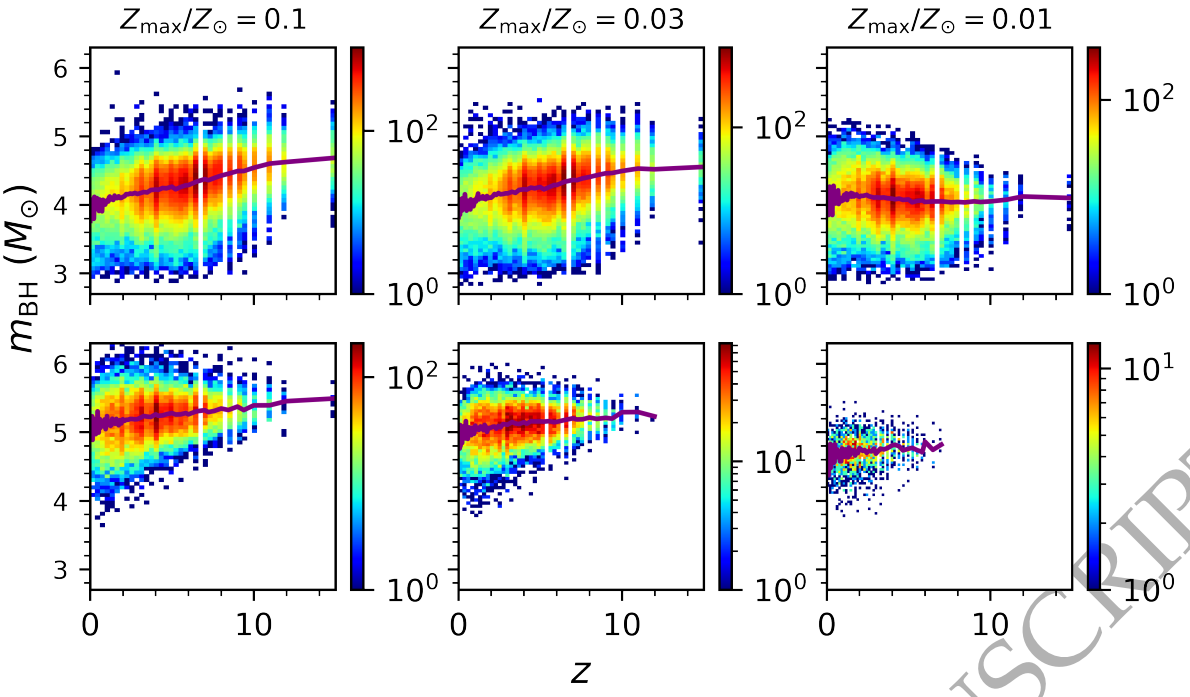

Figure 11. The BH seed mass distributions are shown at the time of seed formation with the color scale showing the number of BH seeds in each pixel of the redshift - BH mass plane. This is done for the model variations where $m_{\rm gas,min} = 10^8 M_\odot$ (top panel) and $m_{\rm gas,min} = 10^9 M_\odot$ (bottom panel) with their metallicity variations from $Z_{\rm max} = 0.1$ to 0.03 to 0.01 shown from the left to right panels, respectively. Their corresponding median BH seed masses at each redshift are shown by the purple lines. The median BH seed formation mass is $\sim 10^4 M_\odot$ for $m_{\rm gas,min} = 10^8 M_\odot$ model realizations versus $\sim 10^5 M_\odot$ for $m_{\rm gas,min} = 10^9 M_\odot$. Overall, BH seeding peaks between $z \sim 2-8$.

alizations produce local BHMFs within the range of empirical estimates. Similar to the mass density evolution, the above results are also not surprising given that 1) the BH masses are assigned to scale with the stellar mass consistent with the local $M_* - M_{\rm bh}$ relation with scatter and 2) TNG stellar mass functions are consistent with the observational constraints.

3.5 BH Occupation Fractions

Figure 10 shows the $z = 0-5$ BH occupation fractions ($f_{\rm occ}$) of all the MP model realizations where $m_{\rm gas,min} = 10^8 M_\odot$ (top panel) and $m_{\rm gas,min} = 10^9 M_\odot$ (bottom panel) with their metallicity variations shown in the same previous color scheme and their $f_{\rm seed} = 1$ (solid lines) versus 0.01 (dashed lines) variation results. At each redshift, $f_{\rm occ}$ is defined as the fraction of all halos in a given stellar mass bin that contain a BH. At $z = 0$, all model realizations have $f_{\rm occ} = 1$ in the highest mass bin. In the `mgas8_Z0.01` and `mgas9_Z0.03` realizations, the $z = 0$ occupation fraction drops below 1 at $M_* \sim 10^{10.5} M_\odot$, while for the most lenient seeding criteria, $f_{\rm occ} = 1$ down to $M_* < 10^{8.5} M_\odot$.

At $z > 0$, the BH occupation fraction shows stronger dependence on the seeding criteria, including $f_{\rm seed}$. The $f_{\rm occ}$ results produced by the $Z_{\rm max} = 0.1$ and $0.03 Z_\odot$ variations of the model with $m_{\rm gas,min} = 10^8 M_\odot$ are nearly identical below below $z \sim 3$, just like their mass densities. The `mgas9_Z0.01` model realization, which does not begin seed formation until $z \lesssim 7$ (see Figures 7-8), produces unphysically low occupation fractions across most of the stellar mass and redshift range shown. Note that at $z = 2$, $f_{\rm occ} \to 0$ in the highest stellar mass bin owing to a single massive halo that becomes metal-enriched at very early times (thus preventing seed formation in our model) and then has a quiet merger history until $z < 2$.Note also that numerous previous studies have found that BH occupation fraction depends strongly on the seeding efficiency (Sesana et al. 2007; Volonteri et al. 2008; DeGraf & Sijacki 2020; Spinoso et al. 2023). In our definition of $f_{\rm seed}$, applied at every snapshot, the cumulative probability that a given galaxy will eventually be seeded with a BH is much higher than the value of $f_{\rm seed}$; recall that we do not impose a minimum redshift for seeding. In addition, the impact of $f_{\rm seed}$ at $z = 0$ is further reduced as galaxies merge over time, driving the occupation fraction higher, although mergers will have less of an effect for low-mass halos.

In Figure 10, our model results are also compared with $f_{\rm occ}$ from a sample of dwarf host galaxies chosen from the RomulusC simulation Tremmel et al. (2019) in Tremmel et al. (2024) (hereafter called T24). Also shown is an empirical estimate of the $z = 0$ $f_{\rm occ}$ derived from the optically selected early-type galaxy sample in Miller et al. (2015) (hereafter called M15; 1σ confidence intervals shown for their "clean" sample, which accounts for intrinsic scatter, measurement uncertainties, and X-ray limits.) The `mgas8_Z0.01` and `mgas9_Z0.03` model realizations, which were highlighted above as producing the best overall agreement with other empirical constraints, overlap with the M15 occupation fractions and the T24 simulation predictions over most of the stellar mass range at $z = 0$. At $z = 1$, `mgas8_Z0.01` agrees

very well with the $z = 1$ occupation fractions from T24, while `mgas9_Z0.03` predicts slightly lower $f_{\rm occ}$ at intermediate stellar masses. At $z > 1$, these model realizations have significantly lower $f_{\rm occ}$ than T24 especially at intermediate stellar masses. Meanwhile, our most lenient metallicity criterion ($Z_{\rm max} > 0.1 Z_{\odot}$) yields occupation fractions at or above the M15 and T24 predictions. Note that occupation fractions will be affected by any BH ejections via gravitational-wave recoil or slingshot kicks from three-body interactions, especially at higher redshift (e.g., Volonteri et al. 2010; Blecha et al. 2016; Izquierdo-Villalba et al. 2020; Askar et al. 2021). BH ejections are not accounted for in our model, which may contribute to the discrepancy. Another complicating factor is the difficulty in making direct comparisons between model BHOFs and those derived from EM observations, since we do not model the observational counterparts and Miller et al. (2015) directly probe only active BHs.

The BH occupation fractions provide a key illustration of how our TNG-based SAM is able to populate BHs in more low- to intermediate-mass galaxies than TNG. The TNG occupation fraction at $z = 0$ lies within the range of the M15 predictions (albeit at the lower end of this range). The `mgas8_Z0.01` and `mgas9_Z0.03` model realizations agree with the M15 results for galaxies with $10^8 \lesssim M_*/M_{\odot} \lesssim 10^{9.5-10}$, while the BHOFs for the `mgas8_Z0.03`, `mgas8_Z0.1`, and `mgas9_Z0.1` model realizations skirt the upper edge of the M15 interval in the $M_*/M_{\odot} \sim 10^{9-10}$ mass range.

In Section 3.3.6, we highlighted `mgas8_Z0.01` and `mgas9_Z0.03` as the model realizations that agreed best with the empirical constraints on mass and number density. We conclude this subsection by considering together the BH number density, BH mass density, $z = 0$ BHMF, and BHOF results and comparing them for these two realizations. Both realizations predict higher number and mass densities than those inferred from observations, with slightly higher densities for `mgas8_Z0.01`. As noted above, however, the MP includes BHs down to $10^5 M_{\odot}$, which are not easily probed with observations. BHMFs agree well with observations, with little difference between model realizations (a result of pinning BH masses to a stellar mass scaling relation). BHOFs for both realizations are lower than those predicted by T24 at $z \geq 2$, but at $z = 0 - 1$, the `mgas8_Z0.01` realization agrees quite well with the T24 predictions. The `mgas9_Z0.03` realization matches the T24 predictions at low mass for $z = 0-1$ but underpredicts occupation fractions at high mass. Compared to the empirically derived $z = 0$ BHOF from M15, `mgas8_Z0.01` provides a slightly better match than `mgas9_Z0.03`, though both have a shallower increase in BHOF with stellar mass than that measured by M15.

Thus, none of the model realizations provide a perfect match to all of the empirical constraints considered here. This is undoubtedly a combination of observational biases (such as the sample incompleteness of BHs below $10^6 M_{\odot}$) and model assumptions & uncertainties, many examples of which are discussed elsewhere in the text. Given these uncertainties, the `mgas8_Z0.01` and `mgas9_Z0.03` model realizations produce fairly good agreement with most of the empirical constraints, and they provide a closer match than the other model realizations considered here. Because the `mgas8_Z0.01` produces somewhat better agreement with other constraints on the BHOF, and because the number densities (which `mgas8_Z0.01` overpredicts slightly more than `mgas9_Z0.03`) are more likely to be subject to observational sample incompleteness than other quantities, we deem the `mgas8_Z0.01` model realization to be the best overall match to empirical constraints, among the model realizations considered in this work.

3.6 Newly-formed BH seeds in our TNG SAMs

Here we examine in more detail the properties of BH seeds at the time of formation Figure 11 tracks the median BH seed masses (purple lines) across cosmic time for all model realizations where $m_{\rm gas,min} = 10^8 M_{\odot}$ (top panel) and $m_{\rm gas,min} = 10^9 M_{\odot}$ (bottom panel), with maximum metallicities of $Z_{\rm max} = 0.1$, 0.03, & 0.01. The BH seed mass distribution and color scale correspond to the number of BH seeds in each pixel of the redshift - BH mass plane. This seeding parameter space is highlighted for the same aforementioned reasons as for Figure 10 because it includes `mgas8_Z0.1` and `mgas9_Z0.03`.

Seeding peaks around $z = 2-8$ in all the model realizations for all metallicity thresholds, and seed formation begins in the $z = 15$ or $z = 12$ snapshots for most model realizations. Note that although we do not impose a minimum redshift for seed formation, seeding naturally tapers off below $z \sim 2-3$, as galaxies become increasingly metal-enriched. Comparing Figure 11 with the BH number density evolution (Figures 6 & 7), we see that the slow-down of new seed formation occurs at a similar time as, or slightly later than, the peak in BH number density. This highlights the important role of mergers in galaxy and BH evolution at cosmic noon. Because seeding does persist to $z = 0$ with median seed masses of $\sim 10^{4-5}$, albeit at a slower rate, our assumption of no minimum redshift for BH seeding contributes to the higher mass and number densities in many of our models (relative to empirical constraints).

The median seed mass varies from $\sim 10^4$ to $10^{4.6}$ $M_{\odot}$ in the $m_{\rm gas,min} = 10^8 M_{\odot}$ model realizations, with a mild dependence on metallicity criterion and redshift. The $m_{\rm gas,min} = 10^9 M_{\odot}$ model realizations have higher seed masses overall, as expected, with median masses of $\sim 10^{5-5.5} M_{\odot}$ for the most lenient metallicity criterion and $\sim 10^{4.6-4.8}$ for $Z_{\rm max} = 0.01 Z_{\odot}$. We again emphasize that the BH masses in our model rely on a simple prescription in which the BH mass follows a scaling relation with stellar mass; they must therefore be interpreted with some caution.

The first BH seeds form in the $z = 15$ snapshot in most models in Figure 11, except the `mgas9_Z0.03` realization for which seeding begins at $z = 12$ (and the strictest `mgas9_Z0.01` model that fails to produce a realistic BH population at all). The onset of seeding in our model is therefore consistent with predicted formation redshifts for intermediate-mass and massive BH seeds in many previous works with a variety of different approaches to BH seeding (e.g., Lodato & Natarajan 2006; Davies et al. 2011; Agarwal et al. 2012; Ryu et al. 2016).

In general, BHs are seeded with a smaller range of masses at higher redshift, reflecting the relatively narrow range of host galaxies that simultaneously meet the mass and metallicity criteria. (This is less true for the most lenient seeding criteria.) As time goes on, galaxies with a wider range of evolutionary histories, and thus a wider range of stellar masses, are able to be seeded with BHs. This includes cases in which

galaxies lose stellar mass to tidal stripping and are seeded with slightly lower BH masses at later times.

Similarly to our findings, Agarwal et al. (2012) and Dijkstra et al. (2014) find that the number of seeds formed in their models steadily increases through the epoch of cosmic dawn ($z \sim 6$ and $z \sim 10$, respectively), with a wide range in DCBH number densities depending on factors including critical LW flux and galactic outflows. These works do not consider seed formation at lower redshifts. We can also compare our findings to those of Bonoli et al. (2014) and Spinoso et al. (2023), both of which consider multi-channel BH seed formation and find that the number density of massive seeds is well below the observed number density, such that the descendants of light seeds play a key role in producing the observed BH population. In Spinoso et al. (2023), seed formation does not continue below $z \sim 6$ owing to enrichment of the intergalactic medium. Bonoli et al. (2014) consider a model in which massive seeds are formed during major, gas-rich galaxy mergers, finding that most major mergers meet the conditions for massive seed formation at $z > 4$, and that these massive seeds dominate the massive end of the BHMF at early times. Interestingly, as in our model, their massive seed formation channel also persists to $z \sim 0$ in some systems. However, we caution that a comparison between our results and those of Bonoli et al. (2014) is not straightforward, given the very different approaches to BH seeding.

4 CONCLUSIONS

In this work, we build a novel semi-analytic BH seed model that forms BHs and traces their evolution along galaxy merger trees within the TNG50 volume of the IllustrisTNG simulation suite. We systematically explore a wide range of criteria for seeding a BH in TNG halos. We consider model realizations that seed a BH in each halo that exceeds minimum thresholds in gas mass ($m_{\rm gas,min} = 10^7 - 10^9\ M_\odot$) and total mass ($m_{\rm tot,min} = 10^{8.5} - 10^{10.5}\ M_\odot$), with gas metallicities less than a maximum limit ($Z_{\rm max} = 0.1, 0.03, 0.01\ Z_\odot$). We treat the BHs in our model independently from those in TNG, and we also make the simplifying assumption that at most one BH is present in each halo (i.e., we consider only the total BH mass per halo). The model is motivated by the expectation that popular theoretical seeding channels such as Pop III, NSC and DCBH seeds form in halos with low metallicity and dense (star-forming) gas. We consider intermediate-mass and massive seeds but remain agnostic as to the exact seeding channel. Seeding criteria are based on halo-averaged quantities rather than properties of individual gas cells, and we focus on a regime of parameter space that is well-converged with resolution. This produces a seeding prescription that can be readily applied to larger-volume simulations at lower resolution. The halo mass cuts ensure that the seeding takes place regions with deep enough gravitational potentials and that no seeds form in spuriously identified gas clumps outside of dark matter halos. The gas mass cuts ensure that there is sufficient gas in the halo, a small fraction of which is presumed to actually form the BH seed. To account for the possibility that additional criteria may be required to form BH seeds, we also consider model realizations in which each halo that meets all other criteria forms a BH seed with probability $f_{\rm seed} = 0.01$. Lastly, we also ensure that the seeded halos have at least one star particle, to ensure that these halos have a prior history of assembling dense star forming gas, and because we assign BH masses based on a simple scaling with the host stellar mass.

We first validated our approach by using the original TNG50 seeding criterion in our semi-analytic framework (i.e., seeding BHs in $m_{\rm tot} > 5 \times 10^{10}\ M_\odot h^{-1}$ halos). When these BHs are populated in our halo merger trees, we find that the resulting BH counts are consistent with the BH population produced in the original TNG50 run to within 4% at $z = 0$. We then proceed to make predictions of BH populations for a range of seeding criteria and compare them to empirical constraints from AGN observations (M08, S09, C10, and S20).

Here we highlight our main conclusions:

- A wide range of seeding criteria produce number densities of massive BHs ($> 10^5\ M_\odot$) that are broadly comparable to current empirical measurements. Only one of our strictest model realizations (`mtot10.5_Z0.01`) completely fails to produce enough BHs at any epoch. The most lenient seeding criteria produce somewhat more BHs than the TNG simulations as well as empirical measurements at $z \sim 0$, with the exception of S09. However, note that there is uncertainty among the empirical measurements at $z \sim 0$, with very few constraints at the low-mass end ($\sim 10^5 - 10^6\ M_\odot$, which S09 includes). At higher redshifts, the empirical constraints are even more uncertain. Most of our model realizations predict higher number densities than these measurements, especially at high redshift. This tension reflects the large population of low-mass BHs in our model, and the dearth of empirical data on this population. Consider also that the TNG scaling relation, and therefore our model, generally overpredicts mass densities before cosmic noon.
- Just as the massive BH populations in our model are dominated by BHs at the low-mass end ($\sim 10^5 - 10^6\ M_\odot$), when we consider the full population of BHs in our model (down to $\sim 10^3\ M_\odot$), we find that the BH number densities are dominated by low-mass ($\sim 10^3 - 10^5\ M_\odot$) BHs. This low-mass population is also more sensitive to changes in the halo or gas mass seeding thresholds. These $< 10^5\ M_\odot$ BHs would be difficult to detect with EM observations, but mergers between them would in many cases be observable with LISA, LIGO-Virgo-KAGRA, and next-generation ground-based GW detectors. We will quantify massive BH merger rates for our model in forthcoming work.
- Much less variation is seen in the BH mass densities, all of which converge to a narrow range of values at $0 \lesssim z \lesssim 4$ consistent with empirical estimates (this excludes the aforementioned strictest `mtot10.5_Z0.01` seed model). The good agreement in mass densities is a natural consequence of our BH mass growth model in which BH masses simply trace the host stellar mass, given the success of TNG simulations in reproducing the observational constraints for the galaxy stellar mass function and cosmic star formation rate density. However, at higher redshifts ($z \gtrsim 4$), our seed model realizations start to diverge in their mass density predictions for the massive $> 10^5\ M_\odot$ BHs (up to nearly 3.5 orders in magnitude). At these redshifts, it is the low mass BHs that dominate the BH mass density, particularly for the more lenient seeding criteria. This underscores the importance of LISA for the potential detection of these low mass BHs to constrain the

high-z BH mass density, and hence the underlying seeding channels.

- Our BHMFs are very similar to the TNG BHMFs at the high-mass end ($\gtrsim 10^9\ M_\odot$), but our model BHMFs are consistently lower than those in TNG at lower masses. This is also reflected in the slightly lower BH mass densities relative to TNG, which seeds only massive BHs ($8 \times 10^5\ M_\odot h^{-1}$). Both our BHMFs and the TNG BHMF fall within the range of empirical measurements for the majority of our seed model realizations. Again, these comments exclude the strictest `mtot10.5_Z0.01` seed model realization that produces too few seeds. Additionally, the `mtot10.5_Z0.03` and `mgas9_Z0.03` model realizations also somewhat under-produce the $\lesssim 10^8$ $M_\odot$ BHs. At the other end, the more lenient seeding criteria produce nearly identical $z = 0$ BHMFs, which reflects their consistent $z = 0$ BH occupation fraction of essentially unity for halos resolved in TNG.
- The $z = 0$ BH occupation fractions produced in our model are in good agreement with empirical constraints (M15) for model realizations that best agree with observed BHMFs, mass, and number densities. In particular, these model realizations produce higher BH occupation fractions than TNG in low-mass and intermediate-mass galaxies, highlighting the ability of our model to yield more realistic BH populations than TNG especially in dwarf galaxies than TNG.
- Considering together the BH mass and number densities, BHMFs, occupation fractions, and new seed masses, we highlight `mgas8_Z0.01` and `mgas9_Z0.03` as model realizations that produce the best agreement with empirical data, including the requirement that seed formation begins before the epoch of the first quasars. The seed masses in these model realizations range from $\sim 10^3 - 10^6 M_\odot$, with medians between $\sim 10^4$ and $10^{5.5} M_\odot$. We deem the `mgas8_Z0.01` realization to be a slightly better overall match than `mgas9_Z0.03`, owing primarily to its somewhat better agreement with literature constraints on BH occupation fractions.
- We still find that several seeding variations produce BH populations in reasonable agreement with observations. In most cases, our model realizations do produce higher BH number densities than those inferred from empirical data, but this is significantly influenced by the inclusion of $10^5 - 10^6$ $M_\odot$ BHs in our massive BH population, a mass regime where few empirical constraints exist. We note that a combination of the varying-$m_{\rm gas,min}$ and varying-$m_{\rm tot,min}$ cuts produces similar results to those presented here. In several cases, reasonable $z = 0$ BH populations are produced when combining these mass cuts with a maximum gas metallicity ranging from $0.01 - 0.1\ Z_\odot$ and a seeding probability from $0.01 - 1$. The exception is the strictest metallicity cut ($Z_{\rm max} < 10^{-2}\ Z_\odot$) combined with the strictest mass cuts ($m_{\rm tot} > 10^{10.5}\ M_\odot$ or $m_{\rm gas,min} = 10^9\ M_\odot$); these model realizations produce few if any BHs at $z > 6$ and cannot reproduce the $z = 0$ BH population.

Until the BHMF and its redshift evolution are more well-determined, this uncertainty will continue to be a barrier for models of BH formation and evolution, particularly in the low-mass and high-redshift regimes. JWST is pushing the envelope, being able to observe both bright and faint quasars earlier than previously possible (Larson et al. 2023; Onoue et al. 2023; Scholtz et al. 2023; Fujimoto et al. 2022). Paired with GW observations of SMBH binaries expected from LISA as far back $z \approx 20$, this will greatly increase our understanding of BH populations at early cosmic times. In turn, these data will constrain theoretical models of BH formation and early evolution, allowing us to probe the elusive origins of massive BHs.

ACKNOWLEDGEMENTS

AE, LB, & AB acknowledge support from NSF awards AST-1909933 and AST-2307171, and LB acknowledges support from the Research Corporation for Science Advancement under Cottrell Scholar Award #27553. We also thank Paul Torrey and Luke Kelley for helpful discussions on the results.

DATA AVAILABILITY

Data from the sublink and merger tree catalogs for the TNG50 simulation used in this project may be found on the TNG Project website: https://www.tng-project.org/data/.

Scripts that retrieve descendants from the TNG merger trees may be found in this Github repository: https://github.com/akbhowmi/arepo_package.

REFERENCES

Abbott B. P., et al., 2016, Phys. Rev. D, 93, 122003
Ade P. A. R., et al., 2016, Astronomy & Astrophysics, 594, A13
Agarwal B., Khochfar S., Johnson J. L., Neistein E., Dalla Vecchia C., Livio M., 2012, MNRAS, 425, 2854
Agarwal B., Regan J., Klessen R. S., Downes T. P., Zackrisson E., 2017, MNRAS, 470, 4034
Agazie G., et al., 2023a, ApJ, 951, L8
Agazie G., et al., 2023b, ApJ, 952, L37
Ahn E.-J., Bertone G., Merritt D., Zhang P., 2007, Phys. Rev. D, 76, 023517
Amaro-Seoane P., et al., 2017, arXiv e-prints, p. arXiv:1702.00786
Amaro-Seoane P., et al., 2023, Living Reviews in Relativity, 26, 2
Antoniadis J., et al., 2023, arXiv e-prints, p. arXiv:2306.16214
Askar A., Davies M. B., Church R. P., 2021, MNRAS, 502, 2682
Askar A., Davies M. B., Church R. P., 2022, MNRAS, 511, 2631
Bañados E., et al., 2018, ApJ, 856, L25
Barausse E., 2012, MNRAS, 423, 2533
Becerra F., Greif T. H., Springel V., Hernquist L. E., 2015, MNRAS, 446, 2380
Begelman M. C., Silk J., 2023, arXiv e-prints, p. arXiv:2305.19081
Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
Bhowmick A. K., et al., 2021, MNRAS, 507, 2012
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
Bhowmick A. K., Blecha L., Torrey P., Weinberger R., Kelley L. Z., Vogelsberger M., Hernquist L., Somerville R. S., 2024, MNRAS, 529, 3768
Blecha L., Loeb A., 2008, MNRAS, 390, 1311
Blecha L., et al., 2016, MNRAS, 456, 961
Boekholt T. C. N., Schleicher D. R. G., Fellhauer M., Klessen R. S., Reinoso B., Stutz A. M., Haemmerlé L., 2018, MNRAS, 476, 366
Bond J. R., Arnett W. D., Carr B. J., 1984, ApJ, 280, 825
Bonoli S., Mayer L., Callegari S., 2014, MNRAS, 437, 1576

Booth C. M., Schaye J., 2011, MNRAS, 413, 1158
Bromm V., Larson R. B., 2004, ARA&A, 42, 79
Bromm V., Loeb A., 2003, ApJ, 596, 34
Bromm V., Yoshida N., 2011, ARA&A, 49, 373
Buchner J., Treister E., Bauer F. E., Sartori L. F., Schawinski K., 2019, ApJ, 874, 117
Cao X., 2010, ApJ, 725, 388
Chabrier G., 2003, PASP, 115, 763
Chon S., Hosokawa T., Omukai K., 2021, MNRAS, 502, 700
Clark P. C., Bonnell I. A., Klessen R. S., 2008, MNRAS, 386, 3
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Davies M. B., Miller M. C., Bellovary J. M., 2011, ApJ, 740, L42
DeGraf C., Sijacki D., 2020, MNRAS, 491, 4973
Decarli R., et al., 2012, ApJ, 756, 150
Dijkstra M., Haiman Z., Mesinger A., Wyithe J. S. B., 2008, MNRAS, 391, 1961
Dijkstra M., Ferrara A., Mesinger A., 2014, MNRAS, 442, 2036
Dubois Y., et al., 2014, MNRAS, 444, 1453
Fan X., Carilli C. L., Keating B., 2006, ARA&A, 44, 415
Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
Fluetsch A., et al., 2019, MNRAS, 483, 4586
Fragione G., Silk J., 2020, MNRAS, 498, 4591
Fujimoto S., et al., 2022, Nature, 604, 261
Ge Q., Wise J. H., 2017, MNRAS, 472, 2773
Genel S., et al., 2018, MNRAS, 474, 3976
Greene J. E., Strader J., Ho L. C., 2020, ARA&A, 58, 257
Habouzit M., et al., 2021, MNRAS, 503, 1940
Habouzit M., et al., 2022, MNRAS, 511, 3751
Harikane Y., et al., 2023, arXiv e-prints, p. arXiv:2303.11946
Hernquist L., Springel V., 2003, MNRAS, 341, 1253
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Inayoshi K., Ichikawa K., 2024, ApJ, 973, L49
Inayoshi K., Omukai K., 2012, MNRAS, 422, 2539
Inayoshi K., Visbal E., Haiman Z., 2020, ARA&A, 58, 27
Izquierdo-Villalba D., Bonoli S., Dotti M., Sesana A., Rosas-Guevara Y., Spinoso D., 2020, MNRAS, 495, 4681
Jahnke K., et al., 2009, ApJ, 706, L215
Jappsen A.-K., Mac Low M.-M., Glover S. C. O., Klessen R. S., Kitsionas S., 2009, ApJ, 694, 1161
Jiang Y., Huang Q.-G., 2022, arXiv e-prints, p. arXiv:2203.11781
Juodžbalis I., et al., 2023, MNRAS, 525, 1353
Kelley L. Z., Blecha L., Hernquist L., 2017, MNRAS, 464, 3131
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
Kocevski D. D., et al., 2023, arXiv e-prints, p. arXiv:2302.00012
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Labbe I., et al., 2023, arXiv e-prints, p. arXiv:2306.07320
Lang R. N., Hughes S. A., 2006, Phys. Rev. D, 74, 122001
Lang R. N., Hughes S. A., 2007, Phys. Rev. D, 75, 089902
Larson R. L., et al., 2023, arXiv e-prints, p. arXiv:2303.08918
Latif M. A., Ferrara A., 2016, Publ. Astron. Soc. Australia, 33, e051
Latif M. A., Whalen D. J., Khochfar S., Herrington N. P., Woods T. E., 2022, Nature, 607, 48
Li Y., et al., 2020, ApJ, 895, 102
Lodato G., Natarajan P., 2006, MNRAS, 371, 1813
Madau P., Rees M. J., 2001, ApJ, 551, L27
Maiolino R., et al., 2023, arXiv e-prints, p. arXiv:2305.12492
Maiolino R., et al., 2024, Nature, 627, 59
Marinacci F., et al., 2018, MNRAS, 480, 5113
Marshall M. A., et al., 2020, ApJ, 900, 21
Martín-Navarro I., Pillepich A., Nelson D., Rodriguez-Gomez V., Donnari M., Hernquist L., Springel V., 2021, Nature, 594, 187
Matthee J., et al., 2023, arXiv e-prints, p. arXiv:2306.05448
Mayer L., Kazantzidis S., Escala A., Callegari S., 2010, Nature, 466, 1082
Mayer L., Fiacconi D., Bonoli S., Quinn T., Roškar R., Shen S., Wadsley J., 2015, ApJ, 810, 51
Mayer L., Capelo P. R., Zwick L., Di Matteo T., 2024, ApJ, 961, 76
McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184
Merchán M. E., Zandivarez A., 2005, ApJ, 630, 759
Merloni A., Heinz S., 2008, MNRAS, 388, 1011
Miller M. C., Hamilton D. P., 2002, MNRAS, 330, 232
Miller B. P., Gallo E., Greene J. E., Kelly B. C., Treu T., Woo J.-H., Baldassare V., 2015, ApJ, 799, 98
Mortlock D. J., et al., 2011, Nature, 474, 616
Naiman J. P., et al., 2018, MNRAS, 477, 1206
Nelson D., et al., 2019a, Computational Astrophysics and Cosmology, 6, 2
Nelson D., et al., 2019b, Monthly Notices of the Royal Astronomical Society, 490, 3234–3261
Neumayer N., Seth A., Böker T., 2020, A&ARv, 28, 4
Omukai K., Schneider R., Haiman Z., 2008, ApJ, 686, 801
Omukai K., Hosokawa T., Yoshida N., 2010, ApJ, 722, 1793
Onoue M., et al., 2023, ApJ, 942, L17
Pacucci F., Loeb A., 2022, MNRAS, 509, 1885
Pacucci F., Loeb A., 2024, ApJ, 964, 154
Pacucci F., Loeb A., Juodžbalis I., 2024, Research Notes of the American Astronomical Society, 8, 105
Pakmor R., Springel V., 2013, MNRAS, 432, 176
Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392
Pillepich A., et al., 2018a, MNRAS, 473, 4077
Pillepich A., et al., 2018b, MNRAS, 475, 648
Press W. H., Davis M., 1982, ApJ, 259, 449
Press W. H., Schechter P., 1974, ApJ, 187, 425
Reardon D. J., et al., 2023, ApJ, 951, L6
Rees M. J., 1984, ARA&A, 22, 471
Regan J. A., Visbal E., Wise J. H., Haiman Z., Johansson P. H., Bryan G. L., 2017, Nature Astronomy, 1, 0075
Reines A. E., Volonteri M., 2015, ApJ, 813, 82
Ricarte A., Natarajan P., 2018, MNRAS, 481, 3278
Rodriguez-Gomez V., et al., 2015, MNRAS, 449, 49
Ryu T., Tanaka T. L., Perna R., Haiman Z., 2016, MNRAS, 460, 4122
Sakurai Y., Yoshida N., Fujii M. S., Hirano S., 2017, MNRAS, 472, 1677
Sassano F., Schneider R., Valiante R., Inayoshi K., Chon S., Omukai K., Mayer L., Capelo P. R., 2021, MNRAS, 506, 613
Savorgnan G. A. D., Graham A. W., Marconi A., Sani E., 2016, ApJ, 817, 21
Schaye J., et al., 2015, MNRAS, 446, 521
Schneider R., Omukai K., Inoue A. K., Ferrara A., 2006, MNRAS, 369, 1437
Scholtz J., et al., 2023, arXiv e-prints, p. arXiv:2306.09142
Sesana A., Volonteri M., Haardt F., 2007, MNRAS, 377, 1711
Shankar F., Weinberg D. H., Miralda-Escudé J., 2009, ApJ, 690, 20
Shen X., Hopkins P. F., Faucher-Giguère C.-A., Alexander D. M., Richards G. T., Ross N. P., Hickox R. C., 2020, MNRAS, 495, 3252
Shields G. A., Gebhardt K., Salviander S., Wills B. J., Xie B., Brotherton M. S., Yuan J., Dietrich M., 2003, ApJ, 583, 124
Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575
Smith B. D., Turk M. J., Sigurdsson S., O'Shea B. W., Norman M. L., 2009, ApJ, 691, 441
Spinoso D., Bonoli S., Valiante R., Schneider R., Izquierdo-Villalba D., 2023, MNRAS, 518, 4672
Springel V., 2010, ARA&A, 48, 391
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V., et al., 2018, MNRAS, 475, 676

Suh H., Civano F., Trakhtenbrot B., Shankar F., Hasinger G., Sanders D. B., Allevato V., 2020, ApJ, 889, 32
Tagawa H., Umemura M., Gouda N., 2016, MNRAS, 462, 3812
Tagawa H., Haiman Z., Kocsis B., 2020, ApJ, 892, 36
Taylor P., Kobayashi C., 2014, MNRAS, 442, 2751
Torrey P., et al., 2020, MNRAS, 497, 5292
Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121
Tremmel M., et al., 2019, MNRAS, 483, 3336
Tremmel M., Ricarte A., Natarajan P., Bellovary J., Sharma R., Quinn T. R., 2024, The Open Journal of Astrophysics, 7, 26
Truong N., Pillepich A., Werner N., 2021, MNRAS, 501, 2210
Übler H., et al., 2023, arXiv e-prints, p. arXiv:2302.06647
Valentini M., Gallerani S., Ferrara A., 2021, MNRAS, 507, 1
Valiante R., Schneider R., Zappacosta L., Graziani L., Pezzulli E., Volonteri M., 2018, MNRAS, 476, 407
Vecchio A., Wickham E. D., Stevens I. R., 2004, in Hough J., Sanders G. H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 5500, Gravitational Wave and Particle Astrophysics Detectors. pp 183–193, doi:10.1117/12.551415
Visbal E., Haiman Z., Bryan G. L., 2014, MNRAS, 445, 1056
Vogelsberger M., et al., 2014, Nature, 509, 177
Volonteri M., Natarajan P., 2009, Monthly Notices of the Royal Astronomical Society, 400, 1911–1918
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
Volonteri M., Lodato G., Natarajan P., 2008, MNRAS, 383, 1079
Volonteri M., Gültekin K., Dotti M., 2010, MNRAS, 404, 2143
Wang E. X., Taylor P., Federrath C., Kobayashi C., 2019, MNRAS, 483, 4640
Weinberger R., et al., 2017, MNRAS, 465, 3291
Weinberger R., et al., 2018, MNRAS, 479, 4056
Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32
Wise J. H., Regan J. A., O'Shea B. W., Norman M. L., Downes T. P., Xu H., 2019, Nature, 566, 85
Wyithe J. S. B., Loeb A., 2003, ApJ, 595, 614
Xu H., et al., 2023, Research in Astronomy and Astrophysics, 23, 075024
Yajima H., Khochfar S., 2016, MNRAS, 457, 2423
Zwick L., Mayer L., Haemmerlé L., Klessen R. S., 2023, MNRAS, 518, 2076
van Daalen M. P., Schaye J., 2015, MNRAS, 452, 2247

APPENDIX A: SEEDING IN (CENTRAL) SUBHALOS VERSUS HALOS

In Figure A1, we reproduce one of the best-matching lenient mass cut model realizations with respect to TNG and empirical results, except now with these criteria applied to subhalos, CSHs, and halos, respectively. The MP is shown for SAM realizations where $m_{\rm gas,min} = 10^8\ M_\odot$ and for different values of $Z_{\rm max}$. We see that the differences between seeding in subhalos versus CSHs are extremely small, and the halo-based version of our model produces fairly similar results as well. This is partly because we restrict seeds to one per host, an assumption we discuss in more detail below.

The halo-based version of our model produces slightly more BHs at earlier times and has lower BH number densities at later times, compared to subhalos or CSHs. The halo-based version of our model also has less variation between different metallicity criteria. Note the exact same seeding parameter values have somewhat different physical meanings when applied to halos versus (central) subhalos, so we do not expect the halo-based model version to have identical results to the subhalo- and CSH-based versions of our model. In other words, a central subhalo with a gas mass of $10^8\ M_\odot$ will typically have a halo gas mass larger than this value, so applying the same $m_{\rm gas} > 10^8\ M_\odot$ cut to halos is in effect a stricter criterion than it is for CSHs or subhalos. With this in mind, the level of agreement between these three versions of our model (subhalo-based, CSH-based, and halo-based) provides reassurance that our results do not depend strongly on the type of host used to construct the seeding model.

We also comment on some potential implications of our model assumption of one BH per host. One popular scenario in the literature for massive seed formation via direct collapse involves close pairs of proto-galaxies (possibly within the same DM halo), one of which begins forming stars slightly earlier and is able to irradiate gas in the other galaxy with Lyman-Werner radiation, thereby providing a means for H2 dissociation to delay fragmentation and facilitate monolithic collapse of the gas cloud to a direct-collapse BH (e.g., Dijkstra et al. 2008; Visbal et al. 2014; Regan et al. 2017; Agarwal et al. 2017). By assuming BH formation in CSHs, we are not sensitive to the potential seeding scenario where the CSH begins forming stars first and the massive BH forms in a satellite galaxy. We do not consider this a significant limitation on our results, because BH formation through this channel is expected to be rare (i.e., $\ll$ one per halo). However, recent discoveries of high-redshift AGN from JWST hint that BHs form earlier and more abundantly than previously thought (e.g., Onoue et al. 2023; Maiolino et al. 2023; Larson et al. 2023).

This might be an indication that BH formation in satellites is not as rare as predicted, or it might instead mean that massive seed formation occurs through additional channels such as dynamical heating (e.g., Mayer et al. 2010; Wise et al. 2019; Latif et al. 2022; Zwick et al. 2023).

In addition, our assumption of one BH per host will impact the resulting BH merger rate. Multiple BHs are not allowed to form in the same host, which could cause underestimation of the BH merger rate. On the other hand, previous works find that low-mass BHs at high redshift may spend significant time wandering off-center from their hosts; forcing BHs to merge when their hosts merge would overestimate the BH merger timescale in such cases. Although the BH merger rates are not the focus of this paper, they are very important for LISA. We plan to explore these considerations of BH merger dynamics and merger rates in more detail in future work.

APPENDIX B: RESOLUTION CONVERGENCE AND CHOICE OF SEEDING CRITERIA

In principle, the high resolution of TNG50 reaches the regime of high-redshift atomic cooling halos, in that a $M_{\rm vir} = 10^7 M_\odot$ halo would be marginally resolved with about 20 DM particles. However, the construction of a reliable simulation-based semi-analytic model requires that the salient halo properties for BH seeding (including halo mass, gas mass, and average gas-phase metallicity) are well converged with resolution. This is especially important for the applicability of any such model to larger-volume, lower-resolution simulations, where numerical artifacts from resolution non-convergence would be amplified. Here we examine this question and justify our choice to focus on higher-mass, well resolved halos where the

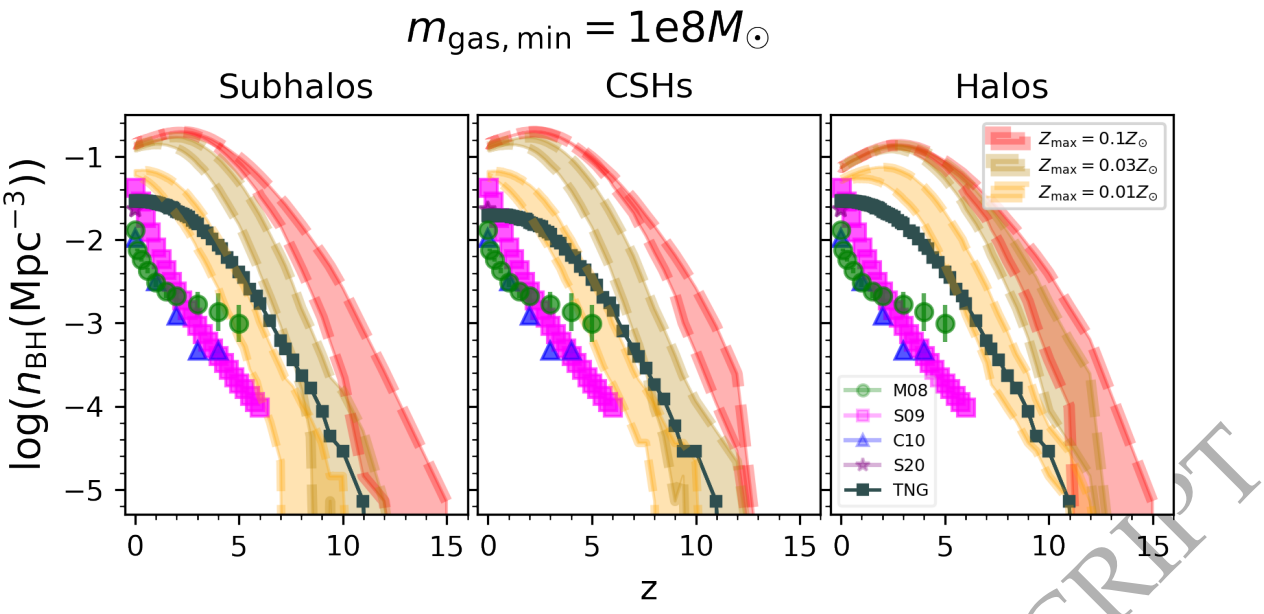

Figure A1. Number density evolution is shown for different hosts in the model realizations where $m_{\rm gas,min} = 10^8\ M_\odot$. There is good agreement between the different host definitions, partly because of the model assumption of one BH per host. The halo-based version of our model seeds BHs relatively earlier than the other hosts and consistently agrees best with TNG and S20 results. These model results highlight the reduction/enhancement of seeding sites between the different hosts. For example, this is naturally a more lenient gas mass cut for halos compared to CSHs, which explains the earlier BH seeding. The halo number densities are lower than in the CSHs at later times, since the group metallicity average is across all subhalo gas cells that belong to that particular halo or galaxy group. There is also less discrepancy between the halo model realizations with different host metallicity criteria, compared to the same corresponding subhalo-based and CSH-based versions of our model.

average gas-phase metallicity (a key seeding criterion in our model) is well converged with resolution.

In Figure B1, we assess the resolution convergence of our results by comparing the BH number density predictions of some of our semi-analytic seed model realizations when applied to three TNG50 volumes with different resolutions. We consider model realizations with $m_{\rm gas,min} = 10^{8.5}\ M_\odot$ and either $m_{\rm tot,min} = 2 \times 10^8 M_\odot$ or $m_{\rm tot,min} = 10^{10}$. The top panels show results for $\log Z_{\rm max}/Z_\odot = -1$ and $\log Z_{\rm max}/Z_\odot = -2$, respectively, while the bottom panel shows results for $\log Z_{\rm max}/Z_\odot = -3$. The $\log Z_{\rm max}/Z_\odot = -1$ case has excellent resolution convergence (with results to within $\sim 0.6-13\%$ between resolutions for each mass cut), while the $\log Z_{\rm max}/Z_\odot = -2$ model realization converges more slowly with increasing resolution. Nonetheless, this case still tends towards convergence. In contrast, the $\log Z_{\rm max}/Z_\odot = -3$ case shows no real trend towards resolution convergence between TNG50-3 and TNG50-1. This occurs because at low resolutions, the average gas-phase metallicity in halos is calculated using a smaller number of gas cells. This limits our ability to resolve the earliest enrichment of halos from their primordial metallicities, and this fact informed our decision not to include metallicity thresholds below $Z_{\rm max} = 0.01\ Z_\odot$ in our seeding model, nor halo mass thresholds below $m_{\rm tot,min} = 10^{8.5}\ M_\odot$.

Our choice of the maximum gas metallicity values for BH seeding follows a similar line of argument. Once a halo begins to form stars, its average gas-phase metallicity evolves very rapidly from the primordial value ($\log(Z_{\rm max}/Z_\odot) = -7$, where many halos reside in Figure 1) to the values probed by our model. Recent simulations with direct, gas-based seeding criteria have found that the formation times of the high-redshift BH population are insensitive to variation of the gas-phase metallicity criterion between $\log(Z_{\rm max} Z_\odot) \sim -3$ to -5 for precisely this reason (Bhowmick et al. 2021, 2022a,b). However, we find that in our model, choosing a metallicity criterion of $\log(Z_{\rm max}/Z_\odot) \lesssim -3$ or -4 would make our results sensitive to the simulation resolution, because even a well-resolved halo may have only a small number of enriched gas cells in this case. Thus, as shown in Figure B1, we do not find acceptable resolution convergence of our results when using these lower values of $Z_{\rm max}$.

We stress that $Z_{\rm max}$ is a maximum gas-phase metallicity (such that halos with lower metallicity will also meet this cut) and that $Z_{\rm max}$ is a halo-averaged quantity (such that a halo with, e.g., $\log Z_{\rm max}/Z_\odot = -2$ could contain metal-poor pockets of gas along with more enriched gas), while a halo with $\log(Z_{\rm max} Z_\odot \lesssim -4)$ would implicitly require the vast majority of the gas content of the seed forming halos to be metal-poor. In previous works with simulations that use the TNG model (Bhowmick et al. 2021, 2022a,b, 2024), most seed-forming halos were not found to require a large fraction of metal-poor gas. In fact, they require only the presence of small pockets of metal-poor gas that are typically surrounded by significantly more metal-enriched regions. This implies that in a seed-forming halo with an average gas-phase metallicity of $\log Z_{\rm max}/Z_\odot = -2$ or -1, the actual seed formation is likely to occur in a local, metal-poor pocket of the halo with a metallicity that is well below the halo average.

We also examine resolution convergence for the subset of BHs in halos with large total mass ($m_{\rm tot,min} = 10^{10}\ M_\odot$),

dashed lines in Figure B1. The convergence is still excellent between TNG50-1 and TNG50-2 for the `mgas8.5_Z0.1` model realization, while the lowest-resolution TNG50-3 simulation somewhat overpredicts number densities at high redshifts. In the `mgas8.5_Z0.01` model realization, the subset of high-mass halos has similar resolution convergence as the full model results, except for TNG50-3, where BH number densities are vastly overpredicted. This indicates that high-mass halos are more susceptible to resolution effects when considering a maximum metallicity threshold, because in low-resolution simulations, enrichment of a few cells might quickly drive the average metallicity of small halos above $Z_{\rm max}$. In contrast, large halos with a few enriched cells are more likely to maintain reservoirs of pristine gas long enough for the BH seeding criteria to be met. Nonetheless, we still find reasonable agreement for the two highest-resolution simulations. In practice, this issue is largely moot in our study; as Figures 7 - 9 show, our low-Z, high-$m_{\rm tot}$ model realizations are the ones that fail to produce a realistic BH population at any redshift.

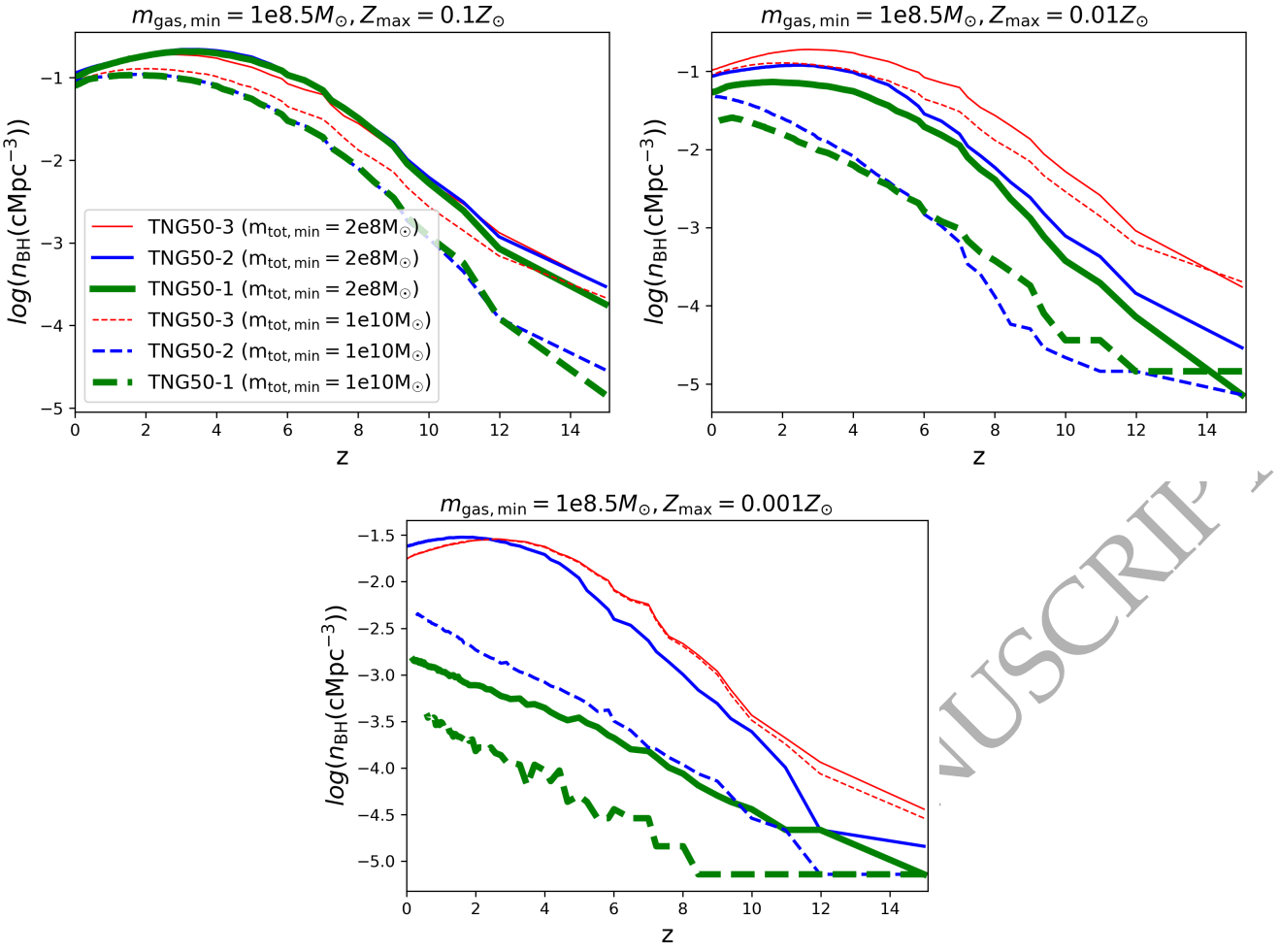

Figure B1. Resolution convergence is quantified through BH number density evolution. In each panel, results are shown for halos with $m_{\rm gas,min} = 10^{8.5}\ M_\odot$ and either $m_{\rm tot,min} = 2 \times 10^8 M_\odot$ (in solid lines) or $m_{\rm tot,min} = 10^{10}\ M_\odot$ (in dashed lines). The panels show results for host metallicity thresholds of $Z_{\rm max}/Z_\odot = 0.1$ (top left), $Z_{\rm max}/Z_\odot = 0.01$ (top right), and $Z_{\rm max}/Z_\odot = 0.001$ (bottom). Model realizations with $Z_{\rm max}/Z_\odot = 0.1$ in the top panel agree between resolutions to within $3-12\%$ and to within $0.6-13\%$ for the lower total mass cut. The average gas-phase metallicity in halos at low resolutions is calculated from fewer gas cells. This makes it difficult to resolve the earliest enrichment of halos from their primordial metallicities. For a given set of mass cuts, lower-metallicity model realizations converge more slowly with resolution, and the $Z_{\rm max}/Z_\odot = 0.001$ shows essentially no trend toward convergence between TNG50-3 and TNG50-1. Therefore, we do not include metallicity thresholds below $Z_{\rm max} = 0.01\ Z_\odot$ in our seeding model.