

PAPER • OPEN ACCESS

Semi-empirical framework of supermassive black hole evolution: highlighting a possible tension between demographics and gravitational wave background

To cite this article: Andrea Lapi *et al* JCAP02(2026)001

View the [article online](#) for updates and enhancements.

You may also like

- [The Implications of High Black Hole Spins for the Origin of Binary Black Hole Mergers](#)
A. Olejak and K. Belczynski
- [Testing the quantum effects near the event horizon with respect to the black hole shadow](#)
Zhaoyi Xu, , Meirong Tang et al.
- [The Observed Mass Distribution of Galactic Black Hole LMXBs Is Biased against Massive Black Holes](#)
Peter G. Jonker, Karamveer Kaur, Nicholas Stone et al.

RECEIVED: July 22, 2025
REVISED: November 22, 2025
ACCEPTED: December 3, 2025
PUBLISHED: February 2, 2026

Semi-empirical framework of supermassive black hole evolution: highlighting a possible tension between demographics and gravitational wave background

Andrea Lapi^{a,b,c,d,*} Francesco Shankar^e Michele Bosi^{a,f} Daniel Roberts^e
Hao Fu^{g,e} Karthik M. Varadarajan^e and Lumen Boco^{h,*}

^aSISSA, Via Bonomea 265, 34136 Trieste, Italy

^bIFPU, Via Beirut 2, 34014 Trieste, Italy

^cINFN — Sezione di Trieste, via Valerio 2, 34127 Trieste, Italy

^dIRA-INAf, Via Gobetti 101, 40129 Bologna, Italy

^eDept. of Physics and Astronomy, Univ. of Southampton, Highfield SO17 1BJ, U.K.

^fDept. of Physics, University of Trento, Via Sommarive 14, 38123 Povo (TN), Italy

^gDept. of Physics, Fudan Univ., Shanghai 200438, P.R. China

^hUniv. Heidelberg, Albert-Ueberle-Str. 3, 69120 Heidelberg, Germany

E-mail: lapi@sissa.it, F.Shankar@soton.ac.uk, mbosi@sissa.it,

d.m.roberts@soton.ac.uk, haofu@fudan.edu.cn,

k.m.varadarajan@soton.ac.uk, lumen.boco@uni-heidelberg.de

ABSTRACT: The evolution of the supermassive Black Hole (BH) population across cosmic times remains a central unresolved issue in modern astrophysics, due to the many noticeable uncertainties in the involved physical processes that span a huge range of spatial, temporal and energy scales. Here we tackle the problem via a semi-empirical approach with minimal assumptions and data-driven inputs. This is based on a continuity plus Smoluchowski equation framework that allows to unitarily describe the two primary modes of BH growth: gas accretion and binary mergers. Key quantities related to the latter processes are incorporated through educated parameterizations, and then constrained in a Bayesian setup from joint observational estimates of the local BH mass function, of the large-scale BH clustering, and of the nano-Hz stochastic gravitational wave (GW) background measured from Pulsar Timing Array (PTA) experiments. We find that the BH accretion-related parameters are strongly dependent on the local BH mass function determination: higher normalizations and flatter high-mass slopes in the latter imply lower radiative efficiencies and mean Eddington ratios with a stronger redshift evolution. Additionally, the binary BH merger rate is estimated to be a fraction $\lesssim 10^{-1}$ of the galaxy merger rate derived from galaxy pairs counts by JWST, and constrained not to exceed the latter at $\gtrsim 2\sigma$. Relatedly, we highlight hints of a possible tension between current

*Corresponding author.

constraints on BH demographics and the interpretation of the nano-Hz GW background as predominantly caused by binary BH mergers. Specifically, we bound the latter's contribution to $\lesssim 30 - 50\%$ at $\sim 3\sigma$, suggesting that either systematics in the datasets considered here have been underestimated so far, or that additional astrophysical/cosmological sources are needed to explain the residual part of the signal measured by PTA experiments.

KEYWORDS: active galactic nuclei, galaxy evolution, gravitational waves / sources, massive black holes

ARXIV EPRINT: [2507.15436](#)

Contents

1	Introduction	1
2	Methodology	4
2.1	The continuity equation	5
2.2	Growth by accretion	6
2.3	Growth by mergers	10
2.4	BH demographics	13
2.5	BH clustering	13
2.6	Stochastic GW background	14
3	Data and analysis	15
3.1	Results and discussion	17
4	Summary	24
A	The high-redshift AGN luminosity function	27
B	Additional tests	27

1 Introduction

Relic black holes (BHs) with masses $M_{\bullet} \sim 10^6 - 10^{10} M_{\odot}$ are found ubiquitously at the center of massive galaxies in the local Universe (e.g., [1–6]). These monsters are thought to have grown mainly by gaseous accretion onto a disk surrounding the BH (e.g., [7, 8]) that energizes the spectacular broadband emission of active galactic nuclei (AGNs) over a broad range of bolometric luminosities $L_{\text{AGN}} \sim 10^{42} - 10^{48} \text{ erg s}^{-1}$. Such a paradigm has received a smoking gun confirmation by the Event Horizon Telescope [9, 10] via the imaging of the BH shadow caused by gravitational light bending and photon capture around the event horizon of M87 and Sgr A*. Nevertheless, the evolutionary history of the supermassive BH population is still an open hot issue of modern astrophysics, as testified by the many studies conducted to address it in the last decades via numerical simulations (e.g., [11–19]), semi-analytic models (e.g., [20–27]), and data-driven approaches (e.g., [28–34]).

In fact, some uncertainties still hinder a robust reconstruction of the growth history of supermassive BHs, even on a population-averaged level. First of all, the initial conditions are poorly known. The simplest picture envisaging light seeds with $M_{\bullet} \lesssim 10^2 M_{\odot}$ of stellar origin growing to the supermassive range by standard disk accretion has been challenged by the discovery of an increasing number of active BHs with masses $M_{\bullet} \gtrsim 10^9 M_{\odot}$ already in place at appreciably high redshifts $z \gtrsim 7$, when the age of the universe was shorter than 0.8 Gyr (e.g., [35–43]; for a recent review see [44]). This may require mechanisms able to rapidly produce heavy BH seeds $M_{\bullet} \sim 10^{3-5} M_{\odot}$, thus reducing the time needed to attain the final masses (see [45–51]).

At the other end, the local demographics of the BH population still suffers from some systematics and possible biases. In fact, large samples of relic, quiescent BHs with precise mass determinations in the local Universe are not available. Thus the local BH mass function (i.e., the comoving number density of BHs as a function of mass) is usually estimated indirectly. A small sample with robust BH mass measurements is exploited to calibrate scaling relations of $M_\bullet$ with other host galaxy properties, like the stellar mass $M_\star$ or the stellar velocity dispersion σ (e.g., [52–57]). Such scaling relations are then employed to convert the statistics of a given galaxy property (e.g., the stellar mass function or the velocity dispersion function, determined on much larger samples of galaxies) into the BH mass function. However, this method is subject to systematics uncertainties that may drive the observed scaling relations away from the intrinsic ones [58–61], hence potentially producing biased estimates of the local BH mass function.

In addition, the intermediate stages of the BH evolution are also somewhat debated. For example, the accretion rates $\dot{M}_\bullet \propto L_{\text{AGN}}/\epsilon$ and radiative efficiencies ϵ for the population of active BHs are loosely constrained, especially toward high-redshift [62–68]. The same applies to the distribution of Eddington ratio $\lambda \propto L_{\text{AGN}}/M_\bullet$ (e.g., [42, 69–73]) or of its proxy $\lambda_\star \propto L_X/M_\star$ involving the X-ray luminosity L_X and host stellar mass $M_\star$ (e.g., [74–77]), which are only vaguely assessed for $z \gtrsim 4$. Even the redshift-dependent AGN luminosity function estimated via multi-wavelength surveys (e.g., [78, 79]) may suffer at $z \gtrsim 6$ from residual systematics on the adopted bolometric corrections and obscured fractions especially at the faint end (e.g., [80–86]). The overall picture has been made even more complex by the recent observations with JWST, which have identified at $z \gtrsim 4.5$ an abundant population of overmassive, X-ray under-luminous BHs (e.g., [87–91]); several of these are located in the so called ‘little red dots’ sources (see [92–96]), whose number density at $z \gtrsim 6$ may be up to two orders of magnitude higher than that of UV/optically selected AGNs [96, 97].

Finally, the role of BH mergers, which constitute another relevant channel for the growth of supermassive BHs, is still under scrutiny. Binary BH merger events are thought to follow (maybe with some delay) the possible aggregation of the host galaxies, common in overdense environments especially toward low redshifts. However, so far such events have remained elusive and difficult to be corroborated, even indirectly, by observations of dual AGNs (e.g., [98–100]). Recently, the situation may have changed with the possible detection of the stochastic gravitational wave (GW) background in the nano-Hz frequency range by pulsar timing array (PTA) experiments [101–104]. Although the statistical significance of the signal does not attain the 5σ threshold of confidence, and superimposed contributions from different physical mechanisms may prevent a clear-cut interpretation, such data offer a new observational handle on the BH growth process via binary mergers, as discussed in the literature with analytic approaches [105–113] and numerical simulations [114–117].

To address the numerous facets and issues of such a complex landscape, three main complementary approaches are possible and have been largely exploited in the last decades: hydrodynamical simulations (for a review, see [118]), semi-analytic models (for a review, see [119]), and semi-empirical frameworks (for a review, see [120]). Hydrodynamical simulations allow to address the galaxy and BH evolution process in fine detail, as they can treat simultaneously dark matter, gas and stars. However, despite the recent increase in resolution

and speed (mainly thanks to emulators and ‘genetic’ codes), many of the relevant physical processes still constitute sub-grid physics, while a detailed exploration of the parameter space is often limited by rather long computational times. Semi-analytic models are based on dark matter merger trees extracted from N -body simulations (or generated via Monte Carlo procedures gauged on them), while the physics inside dark halos is modeled via parametric expressions set on (mainly) local observables. These models are less computationally expensive than hydrodynamical simulations and enable to test different prescriptions more easily, though the considerable number of fudge parameters describing the underlying physics can lead to degenerate solutions and restrict somewhat the predictive power, especially toward high redshift (e.g., [121]). Concerning the evolution of the (super)massive BH population, ab-initio models struggle somewhat in reproducing coherently the wealth of observational data and in elucidating the different role of the main physical mechanisms at work. This is mainly because the problem requires to self-consistently model many diverse processes cooperating on a huge range of temporal, spatial, and energy scales. Such a task goes by far beyond our present physical understanding and computational capabilities, and in any case tends to produce outcomes strongly contingent on the chosen modeling, sub-grid recipes and spatial/temporal resolution (see [122–124]).

The situation calls for a reverse-engineering approach that inverts the traditional methodology: rather than testing ab-initio model predictions against observations, one can construct semi-empirical frameworks that extract physical insights directly from data with minimal theoretical assumptions. While these frameworks may not capture the full theoretical richness of ab-initio models or provide complete physical narratives, their strength lies in delivering robust, model-independent constraints on key observables and providing clear answers to specific, well-posed questions. Moreover, by empirically linking different observables, they can check for possible inconsistencies or tensions among distinct datasets. Such data-driven models serve not to replace more sophisticated semi-analytic and numerical approaches, but are meant as complementary tools to provide independent constraints that can break existing degeneracies and guide theoretical developments toward more physically-motivated parameterizations.

In this work we approach the issue of (super)massive BH evolution via a semi-empirical framework, which is anchored by construction to the most reliable data available so far, takes into account observational/systematic uncertainties by transparently highlighting their impact on the results, and aims to minimize theoretical assumptions and associated parameterizations. Specifically, we employ a continuity plus Smoluchowski equation framework to unitarily describe the two main modes of growth for the supermassive BH population, namely gas accretion and binary mergers. We educatedly parameterize the main uncertainties associated to each BH growth channel, and infer joint constraints on these in a Bayesian setup from estimates of the local BH demographics, clustering, and measurements of the nano-Hz stochastic GW background. The plan of the paper is as follows: in section 2 we describe the ingredients and methodology underlying our semi-empirical framework; in section 3 we present our data analysis and discuss our results; in section 4 we summarize our findings. Throughout this work we adopt the standard Λ CDM cosmology with parameters from the Planck mission [125].

2 Methodology

In this section we first provide a qualitative overview of our data-driven methodology, which is also schematically illustrated in figure 1, while formal details and references can be found in the following subsections.

- The core tool of the method is constituted by the continuity equation. This provides an effective way of deriving the statistical evolution of the (super)massive BH population from data-driven inputs. Specifically, the continuity equation is a partial differential equation which evolves timestep by timestep the BH mass function across cosmic time, taking into account the averaged growth of the BH population due to gas accretion and mergers.
- The component related to gas accretion is described by a drift term in the continuity equation. It requires two *inputs*: the observed redshift-dependent bolometric AGN luminosity function $N(\log L_{\text{AGN}}, t)$, which provides the statistics of active BHs emitting at a given luminosity L_{AGN} ; a parameterized form of the Eddington ratio distribution $P(\log \lambda)$, which basically informs on how BHs of different masses are contributing to the different luminosity bins. The *parameters* involved are three: two ($\log \lambda_0$ and ξ) describe the local normalization and redshift evolution for the mean of the Eddington ratio distribution; the other one is the radiative efficiency ϵ , which allows to convert AGN luminosity into BH accretion rates, eventually needed to compute BH masses.
- The component related to mergers is described by a Smoluchowski coagulation (kind of diffusion) term in the continuity equation. It requires one *input* and two additional *parameters*: in the reasonable assumption that BH mergers follow galaxy mergers, one parameter $f_{\bullet\bullet\rightarrow\bullet}$ is a rescaling factor for the redshift-dependent normalization of the BH merger rate with respect to the galaxy merger rate as estimated from galaxy pair counts, which constitutes the observational input (but we will explore variants where the BH merger rate is maximized by adopting $f_{\bullet\bullet\rightarrow\bullet} \approx 1$); another parameter η specifies a possible additional dependence of the BH binary merger rate on the BH mass.
- The *direct output* of the continuity equation is a unique solution for the amplitude, shape and evolution of the BH mass function $N(\log M_{\bullet}, t)$ as built by accretion and mergers. The BH mass function at $z \approx 0$ can be directly compared with independent observational determinations from local galaxy statistics and BH-to-galaxy scaling relations.
- A first *derived outcome* is the BH clustering. We compute the mean relation between BH mass and host dark matter halo mass via abundance matching of the BH mass function yielded by the continuity equation and the halo mass function $N(\log M_{\text{H}}, t)$ as provided by N -body simulations in the standard Λ CDM cosmology. We then use this relation to convert halo bias into BH bias (i.e., a clustering signal). This procedure requires an additional *parameter*, which is the scatter $\sigma_{\log M_{\text{H}}}$ in the relationship between BH and host halo mass. The BH bias so obtained can be directly compared with observational estimates of BH clustering.

- Another *derived outcome* is the stochastic GW background from BH mergers in the nano-Hz frequency range. This can be straightforwardly computed from the outputs of the continuity equation and in particular from the redshift-dependent BH mass function and binary BH merger rate (related to the Smoluchowski coagulation term). The resulting GW background amplitude can then be directly compared with the current observational estimates from PTA experiments.
- The aforementioned parameters of our semi-empirical framework (namely, the radiative efficiency of BH accretion, the mean of the Eddington ratio distribution and its redshift evolution, the normalization and mass dependence of the BH merger rate, and the scatter in the BH vs. halo mass relation) are *inferred* by comparing the direct output and the derived outcomes with independently measured datasets within a fully Bayesian setup. This approach enables a robust quantification of parameter degeneracies and statistical uncertainties. A key strength of our framework is its ability to jointly constrain the evolution of the supermassive BH population by simultaneously fitting observational data on BH demographics, large-scale clustering, and the GW background. We will discuss below how such a combination of independent observables enables us to highlight a possible tension between the amount of GW background due to binary BH merger and the measured signal from PTA experiments, which could instead be missed by BH merger models that attempt to only fit the GW background ignoring all the other demographic indicators.

In summary, the continuity equation is a flexible framework that by integrating a few observational inputs (e.g., AGN luminosity functions) yields a clear direct output (e.g., BH mass function) and some derived outcomes (e.g., BH clustering, PTA signal) via minimal physical assumptions (e.g., BHs grow by gas accretion and binary mergers) and a limited number of parameters (six in total, see above). The direct output and derived outcomes are then compared with independent datasets to obtain a best estimate of the parameters, which ultimately yields insights into the physical evolution of the BH population. Such a data-driven approach bypasses the complex and uncertain modeling of the baryonic physics associated to galaxies within dark matter halos via semi-analytic recipes or hydrodynamical simulations. Moreover, since the number of observables and parameters are comparable, it is not prone to large degeneracies but only affected by possible systematics. All in all, our semi-empirical framework can robustly assess the relative contributions of accretion and mergers to the evolutionary history of the supermassive BH population, whilst constraining the fraction of the GW background that can plausibly be attributed to supermassive BH mergers.

2.1 The continuity equation

As introduced above, a very effective approach to study the statistical evolution of the supermassive BH population in a data-driven way relies on the continuity equation [29–31, 126–132]. This method allows to self-consistently trace the cosmic evolution of the BH mass function and the associated fraction of active BHs, building on the amount of mass accreted at each cosmic epoch. The latter is extracted from the input AGN luminosity function, from the adopted relation between AGN luminosity and BH mass as specified by the Eddington ratio distribution of active BHs, and from the assumed radiative (plus kinetic)

efficiency. Moreover, being based on a linear differential equation, this approach allows to transparently disentangle the relative roles of accretion and binary mergers as complementary growth channels of the BH population at each cosmic time.

Let $N(M_\bullet, t)$ be the BH mass function, i.e. number density (per comoving volume) of BHs with mass $M_\bullet$ at cosmic time t or at the corresponding redshift z . In the reasonable hypothesis that any individual BH can change mass because of gas accretion and/or of binary merger with a companion, then the evolution of the mass function can be described via the continuity equation:

$$\partial_t N(M_\bullet, t) = -\partial_{M_\bullet} [\langle \dot{M}_\bullet \rangle (M_\bullet, t) N(M_\bullet, t)] + [\partial_t N(M_\bullet, t)]_{\bullet\bullet \rightarrow \bullet} . \quad (2.1)$$

The first addendum on the right hand side is a ‘drift’ term describing growth by gas accretion via a mean accretion rate $\langle \dot{M}_\bullet \rangle$ averaged over the BH population; the underlying rationale is that, although individual BHs may turn on and off, the statistical evolution depends only on the mean accretion rate as a function of mass and time. The second addendum $[\partial_t N(M_\bullet, t)]_{\bullet\bullet \rightarrow \bullet}$ is a ‘coagulation’ term accounting for binary mergers between BHs; note that mergers do not add net mass to the BH population but can alter the mass function by redistribution of the masses, from the low to the high mass end.¹ We now turn to describe these two ingredients in some more detail.

2.2 Growth by accretion

To compute the growth by accretion we rely on one crucial observational input, constituted by the redshift-dependent AGN bolometric luminosity function $N(\log L_{\text{AGN}}, t)$ derived from multi-wavelength data in optical, UV, IR and X-ray bands by [79]. Specifically, we use the analytic rendition by [79] labeled ‘A’ in their table 4. This is expressed as

$$N(\log L_{\text{AGN}}) = \frac{\phi_\star}{(L_{\text{AGN}}/L_\star)^{\gamma_1} + (L_{\text{AGN}}/L_\star)^{\gamma_2}} \quad (2.2)$$

in terms of the parameters $\{\phi_\star(z), \gamma_1(z), \gamma_2(z), L_\star(z)\}$. The latter depend on redshift z according to the expressions

$$\begin{aligned} \log \phi_\star [\text{Mpc}^{-3} \text{dex}^{-1}] &= d_0 + d_1 (1 + z) \\ \gamma_1 &= a_0 + a_1 (1 + z) + a_2 [2 (1 + z)^2 - 1] \\ \gamma_2 &= \frac{2 b_0}{[(1 + z)/3]^{b_1} + [(1 + z)/3]^{b_2}} \\ \log L_\star [L_\odot] &= \frac{2 c_0}{[(1 + z)/3]^{c_1} + [(1 + z)/3]^{c_2}} \end{aligned} \quad (2.3)$$

where $\{a_0, a_1, a_2\} = \{0.8569, -0.2614, 0.02\}$, $\{b_0, b_1, b_2\} = \{2.5375, -1.0425, 1.1201\}$, $\{c_0, c_1, c_2\} = \{13.0088, -0.5759, 0.4554\}$, $\{d_0, d_1\} = \{-3.5426, -0.3936\}$. The resulting bolometric AGN luminosity functions are plotted as a function of redshift over the range $z \sim 0 - 7$ (color-coded) in figure 2.

¹Note, however, that numerically solving the continuity equation on a finite grid of BH masses requires to assume that a constant flux of sub-resolution masses filtrates through the lowest mass boundary, see [133].

We exploit the AGN luminosity function to compute the accretion rate $\langle \dot{M}_\bullet \rangle$ averaged over the population of active BH, which enters in the drift term of the continuity equation (2.1). To make the computation explicit, we need some preliminary definitions and steps. To start with, it is convenient to relate the AGN luminosity L_{AGN} to the BH mass $M_\bullet$. One defines the Eddington ratio $\lambda \equiv L_{\text{AGN}}/L_{\text{Edd}}$, which is the ratio between the actual AGN luminosity and the Eddington luminosity. The Eddington luminosity $L_{\text{Edd}} \equiv 4\pi G \mu_e m_p c M_\bullet / \sigma_T \approx 1.3 \times 10^{38} (M_\bullet/M_\odot) \text{ erg s}^{-1}$ is the limiting value for which the continuum radiation force emitted by the accretion disk around the BH balances gravity in isotropic conditions; in compact form one can write $L_{\text{Edd}} = M_\bullet c^2 / t_{\text{Edd}}$ in terms of the Eddington timescale $t_{\text{Edd}} = c \sigma_T / 4\pi G \mu_e m_p \approx 0.45 \text{ Gyr}$, only dependent on fundamental constants.² Provided these definitions, $L_{\text{AGN}} = \lambda M_\bullet c^2 / t_{\text{Edd}}$ is just a rewriting that by itself does not require any assumption and does not add any relevant information unless one specifies how the Eddington ratio (or the AGN luminosity to BH mass ratio) varies during the lifetime of an individual AGN. However, this rewriting is helpful in view of a statistical description like the one subtended by the continuity equation, because as shown below the demographics (i.e., the mass function) of *active* BHs can be fully characterized by the observed AGN luminosity function and by the redshift-dependent distribution of Eddington ratios.

In fact, as suggested by observations in the local and distant Universe [5, 6, 42, 69, 71–73, 94, 95], active BHs features a distribution of Eddington ratios with a lognormal shape³

$$p(\log \lambda) = \frac{1}{\sqrt{2\pi}\sigma_{\log \lambda}} e^{-(\log \lambda - \langle \log \lambda \rangle)^2 / 2\sigma_{\log \lambda}^2}, \quad (2.4)$$

with dispersion $\sigma_{\log \lambda} \approx 0.35 \text{ dex}$ and with a redshift dependent mean. For the latter we adopt the scaling

$$\langle \log \lambda \rangle(z) = \log \lambda_0 + \xi \log(1+z), \quad (2.5)$$

where λ_0 and ξ are fitting parameters to be determined. We also tried to add a dependence on BH mass, but we found it to be unconstrained with the current data on BH demographics, so we decided to drop it from our analysis.

Then the active BH mass function is linked to the observed AGN luminosity function and to the Eddington ratio distribution by the convolution

$$N(\log L_{\text{AGN}}, t) = \int d \log \lambda p(\log \lambda) N_{\text{AGN}}(\log M_\bullet, t) \Big|_{M_\bullet = L_{\text{AGN}} t_{\text{Edd}} / \lambda c^2}; \quad (2.6)$$

²In this definition, G is the gravitational constant, μ_e is the mean molecular weight per electron, m_p is the proton mass, c is the speed of light, and σ_T is the Thomson cross section.

³We also try to adopt a Schechter-like or a lognormal+powerlaw shape, which have been sometime considered in the literature as inspired from theoretical arguments or empirical analysis of type-2 AGNs at relatively low z (e.g., [28–31, 130, 134]). This requires to add at least one parameter (e.g., powerlaw behavior at small BH masses and its redshift evolution), which however we find to be largely unconstrained given the datasets considered in this work. Thus we prefer to use the pure log-normal expression involving the minimal set of parameters. In any case, we checked that qualitatively our results concerning the buildup of the BH mass function, the role of BH mergers and their contribution to the GW background are unaffected by the choice of the Eddington ratio distribution shape; minor quantitative alterations pertain only to the evolution of the BH-accretion parameters.

this equation may be solved parametrically by assuming a standard double power-law shape⁴ [29, 131, 135] for the active mass function $N_{\text{AGN}} = N_0 / [(M_\bullet / M_{\bullet,0})^{\omega_1} + (M_\bullet / M_{\bullet,0})^{\omega_2}]$ and then by fitting for the best parameters $(N_0, M_{\bullet,0}, \omega_1, \omega_2)(t)$ at any redshift to reproduce the AGN luminosity function. Note that another way to proceed will be to physically model the Eddington ratio distribution by assuming an average light-curve shape for active BHs, as it has been pursued in [28, 30, 130]; however, in the spirit of a data-driven model here we do not take such a route.

To proceed further, we need the BH accretion rate associated to a given AGN luminosity. This is provided by $\dot{M}_\bullet = (1 - \epsilon - \epsilon_{\text{kin}}) L_{\text{AGN}} / \epsilon c^2$ where ϵ and ϵ_{kin} are the radiative and kinetic (jets and/or winds) efficiency, respectively.⁵ Using the previous rewriting of L_{AGN} in terms of BH mass $M_\bullet$, this can also be expressed as $\dot{M}_\bullet = M_\bullet / \tau_{\text{ef}}$ in terms of the e -folding time $\tau_{\text{ef}} \equiv \epsilon t_{\text{Edd}} / \lambda (1 - \epsilon - \epsilon_{\text{kin}})$, which would be the theoretical timescale required for the BH to exponentially grow $M_\bullet \propto e^{\tau / \tau_{\text{ef}}}$ by a factor $e \approx 2.7$ if λ were constant in time. For an average efficiency $\epsilon \sim 0.1$ and $\lambda \sim 1$ it takes on values $\tau_{\text{ef}} \approx 5 \times 10^7$ yr, which is often quoted in textbooks as the typical growth timescale for a BH. However, we stress that such a timescale does not directly correspond to the duty cycle or visibility timescale of an AGN, first because for individual AGNs λ is not constant along the light-curve, second because the emission could be stochastic or strongly intermittent, and finally because the visibility time depends on environmental conditions (e.g., amount of obscuration), on the observational band, and on selection effects (see [136]). For example, literature estimates of the visibility time amount to $\lesssim$ a few 10^7 yr in the optical band for unobscured quasars, while they can be as high as $\gtrsim 10^8$ yr in the hard X-ray band for heavily obscured AGNs [129, 137].

Some discussion on the radiative efficiency ϵ is in order here. For a standard thin disk the radiative efficiency may range from $\epsilon \sim 0.06$ for a non-rotating to $\epsilon \sim 0.42$ for a maximally spinning Kerr BH [64, 138]. However, both for high accretion rates with $\lambda \gtrsim$ a few and for very low accretion rates with $\lambda \lesssim 0.01$, the radiative efficiency tends to decrease due to the onset of an advection dominated regime. In the former case, mainly applying at high redshift, the inflowing gas has a huge density and provides a very large optical depth to trap the radiation and advect it back into the BH [46, 139, 140]. In the latter case, mainly applying to low redshift and in the re-activations of dormant BHs, the inflowing gas density is so low to hinder cooling within an accretion timescale, and to store the viscous energy as a

⁴Although assuming a double power-law shape for the active BH mass function is standard practice, it is worth making a disclaimer here. There is no question that the AGN luminosity function is well described by a double power-law behavior. Since we have assumed a lognormal shape for the Eddington ratio distribution to reflect the latest observational determinations, we are naturally led to adopt a double power-law shape for the active mass function, otherwise the convolution equation (2.8) would not produce the correct shape to fit the luminosity functions. On the other hand, we point out that such an assumption has the direct consequence of yielding a power-law shape in the relic BH mass function at the high-mass end. Despite this is somewhat supported by the data (see section 3.1), the presence of an exponential falloff is still not excluded. Forcing the accretion model to produce it could be achieved by employing a Schechter fit for the active mass function. However, this solution would reduce the high-mass end of the predicted BH mass function at all redshifts, so lowering the BH merger rates and exacerbating the tension with the GW background (see section 3.1).

⁵The radiative efficiency ϵ is conventionally defined with respect to the large-scale mass inflow rate $\dot{M}_{\text{inflow}}$ so that $L_{\text{AGN}} = \epsilon \dot{M}_{\text{inflow}} c^2$, but the actual BH growth rate $\dot{M}_\bullet = (1 - \epsilon - \epsilon_{\text{kin}}) \dot{M}_{\text{inflow}}$ is smaller because of radiative and kinetic losses. Putting together these expressions yields $\dot{M}_\bullet = (1 - \epsilon - \epsilon_{\text{kin}}) L_{\text{AGN}} / \epsilon c^2$.

thermal component that is eventually advected into the hole [141–143]. We have devised a handy λ –dependent expression for the radiative efficiency as a function of the Eddington ratio that describes these various regimes:

$$\epsilon = \epsilon_{\text{thin}} \left[\frac{\lambda/2}{e^{\lambda/2} - 1} - e^{-2(\lambda/\lambda_c)^s} \right], \quad (2.7)$$

with ϵ_{thin} a fitting parameter representing the efficiency of standard disk accretion; the quantities $\lambda_c \approx 0.01$ and $s \approx 1$ determine the behavior in the low accretion regime as suggested by the aforementioned studies, and are held fixed. The above detailed expression is adopted for the sake of completeness, but we anticipate that the advection dominated regimes will turn out to be marginally relevant for the evolution of the BH mass function. This is because the Eddington ratio distribution parameters inferred from our analysis will imply that the thin disk regime is the most typical accretion mode, with mildly super-Eddington values limited to high redshift $z \gtrsim 4 - 6$ and considerably sub-Eddington ones applying only to low $z \lesssim 1$ and large BH masses. As for the kinetic efficiency due to winds and jets, we assume $\epsilon_{\text{kin}} \approx 0.15 \epsilon_{\text{thin}}$ as suggested by [144]; we anticipate that given the average values $\epsilon_{\text{thin}} \lesssim 0.1$ inferred from our analysis, this position implies rather limited kinetic efficiencies. However, note that a few literature studies [145, 146] claim rather higher $\epsilon_{\text{kin}} \lesssim 0.2$, though not applying to all the BH population; in any case, we will explore such a possibility in appendix B. Given the above, the only free parameter of our model specifying the relation between luminosity and accretion rate is the thin-disk radiative efficiency ϵ_{thin} .

We have now all the ingredients to compute the mean accretion rate $\langle \dot{M}_\bullet \rangle$ entering equation (2.1). This is given by

$$\langle \dot{M}_\bullet \rangle(M_\bullet, t) = \frac{M_\bullet}{t_{\text{Edd}}} \int d \log \lambda \, p(\log \lambda) \, \lambda \frac{1 - \epsilon - \epsilon_{\text{kin}}}{\epsilon} \frac{N_{\text{AGN}}(\log M_\bullet, t)}{N(\log M_\bullet, t)}, \quad (2.8)$$

where $p(\log \lambda)$ is the Eddington ratio distribution from equation (2.4), $N_{\text{AGN}}(\log M_\bullet, t)$ is the active BH mass function from equation (2.6) and ϵ is the λ –dependent radiative efficiency from equation (2.7). To better catch the meaning of this expression, note that the quantity $\lambda(1 - \epsilon - \epsilon_{\text{kin}}) M_\bullet / \epsilon t_{\text{Edd}} = M_\bullet / \tau_{\text{ef}} = \dot{M}_\bullet$ is just the accretion rate $\dot{M}_\bullet$ of an individual BH, and the ratio $N_{\text{AGN}}(\log M_\bullet, t) / N(\log M_\bullet, t) \equiv \langle \delta \rangle(M_\bullet, t)$ between the mass function of active and overall BH population defines the so called mean duty cycle $\langle \delta \rangle(M_\bullet, t)$. Thus the above expression can be also written in compact form as $\langle \dot{M}_\bullet \rangle = \int d \log \lambda \, p(\log \lambda) \dot{M}_\bullet \langle \delta \rangle$, highlighting that the underlying rationale is just to weight the accretion rate of a BH by the duty cycle and to average it over the Eddington ratio distribution.

Note that in the continuity equation only the combination $\langle \dot{M}_\bullet \rangle N(\log M_\bullet) = \int d \log \lambda \, p(\log \lambda) \dot{M}_\bullet N_{\text{AGN}}$ enters the drift term describing accretion, and this depends just on the Eddington ratio distribution and the active BH mass function, not on the duty cycle. In fact, in our semi-empirical framework the duty cycle $\langle \delta \rangle(\log M_\bullet, z)$ is not modeled ab-initio and does not serve as an input (apart possibly for the initial condition, see section 2.4), but it is a derived outcome that can be computed after solving the continuity equation for the total BH mass function. However, a detailed comparison of the duty cycle so obtained with observations is not straightforward. First, because observational estimates of the duty cycle

as a function of BH mass are appreciably uncertain, especially toward intermediate to high redshift. Second, because the estimates of the duty cycle depend on the observational band and on selection effects, like the luminosity threshold for an AGN to be considered as active. Nevertheless, we anticipate that in section 3.1 our semi-empirical framework will be shown to produce local mean duty cycles in qualitative agreement with current observations.

It is worth stressing that for pure accretion (no mergers), constant radiative efficiency ϵ (no kinetic output) and Dirac- δ Eddington ratio distribution centered on a mean value $\langle\lambda\rangle$ independent of redshift, the duty cycle is just the ratio of the AGN luminosity and total BH mass function, so that the continuity equation can be readily integrated to yield

$$N(M_\bullet, t) = -\frac{(1-\epsilon)\langle\lambda\rangle^2 c^2}{\epsilon t_{\text{Edd}}^2 \ln(10)} \int dt \left[\partial_{L_{\text{AGN}}} N(\log L_{\text{AGN}}, t) \right] \Big|_{L_{\text{AGN}=\langle\lambda\rangle M_\bullet c^2/t_{\text{Edd}}} . \quad (2.9)$$

which recovers the simplest solution pioneered by [128, 129].

2.3 Growth by mergers

The most general framework⁶ to describe the evolution of the mass distribution $N(M_\bullet, t)$ for a population of BHs under binary aggregations is provided by the Smoluchowski coagulation equation [113, 150–154]:

$$\begin{aligned} [\partial_t N(M_\bullet, t)]_{\bullet\bullet\rightarrow\bullet} = & \frac{1}{2} \int_0^{M_\bullet} dM'_\bullet K(M'_\bullet, M_\bullet - M'_\bullet, t) N(M'_\bullet, t) N(M_\bullet - M'_\bullet, t) \\ & - N(M_\bullet, t) \int_0^\infty dM'_\bullet K(M_\bullet, M'_\bullet, t) N(M'_\bullet, t) , \end{aligned} \quad (2.10)$$

where $K(M_\bullet, M'_\bullet, t)$ is a time-dependent coagulation kernel, which has dimension of inverse of a time (i.e., a merging rate); the kernel is routinely taken to be symmetric in its mass dependencies, i.e. $K(M_\bullet, M'_\bullet, t) = K(M'_\bullet, M_\bullet, t)$. The first integral on the right hand side of the above equation is a production term which refers to a merging event between masses $M'_\bullet$ and $M_\bullet - M'_\bullet$ to produce the mass $M_\bullet$. The factor 1/2 avoids over-counting since the labels of the two merging masses may be interchanged given the symmetric nature of the kernel; it could be removed by changing the upper limit of integration to $M_\bullet/2$. The second addendum is a destruction term which accounts for merging between the mass $M_\bullet$ and a companion with mass $M'_\bullet$.

We stress that the Smoluchowski equation is only appropriate when the population evolution is dominated by binary aggregations, and instead should not be employed when the probability of multiple interactions is appreciable. For example, it is not very apt to describe the merging history of dark matter halos at high redshift (or during any fast accretion phase),

⁶In the literature, the evolution of the BH mass function has been approached with a Fokker-Planck (diffusion-like) equation [29, 147, 148] but this is only an approximation. In fact, it can be shown (for details, see [149]) that the Smoluchowski equation reduces to a Fokker-Planck equation when the mass distribution is sharply peaked around a mean value and the system is in a regime where small changes in mass dominate. These conditions, however, are not guaranteed to hold during the overall evolution of the BH population. Moreover, the Fokker-Planck approximation is not appropriate to distinguish between the production and destruction terms due to binary mergers; an unambiguous assessment of the former is instead required to estimate the GW background.

since the kernels inferred from the excursion set formalism or extracted from numerical simulations imply that the average number of halo mergers is often larger than two (e.g., [152, 155]). However, for supermassive BHs the binary merger assumption is more reasonable. In fact, the issue has been somewhat investigated via aimed dynamical simulations [156, 157]. These show that only a minority around 10 – 20% of binary supermassive BHs can form a triple system by capturing another BH intruder. Moreover, only a fraction $\lesssim 5\%$ actually form a triple systems at parsec scale experiencing chaotic three-body dynamical interaction; in the remaining majority of the cases two of the three BHs merge first and the third is ejected or left orbiting at large separation, so that the merger is still binary. Thus despite merging triple BH systems could be interesting for future GW detector like LISA since they are expected to feature some peculiar signatures, their overall number is very small, and their contribution to the stochastic GW background irrelevant, compared to the standard binary merger channel.

If the mass function evolves under pure coagulation (i.e., no accretion, hence $\partial_t N \equiv [\partial_t N]_{\bullet \bullet \rightarrow \bullet}$) and if the kernel does not rise too steeply as a function of the mass, then by simple algebraic manipulation of equation (2.10) one can write a symmetrized equation for the moments of the mass function in the form

$$\begin{aligned} \partial_t \langle M_{\bullet}^p \rangle &= \partial_t \int_0^\infty dM_{\bullet} N(M_{\bullet}, t) M_{\bullet}^p = \\ &= \int_0^\infty dM'_{\bullet} K(M_{\bullet}, M'_{\bullet}, t) N(M_{\bullet}, t) N(M'_{\bullet}, t) [(M_{\bullet} + M'_{\bullet})^p - M_{\bullet}^p - M'^p_{\bullet}] , \end{aligned} \quad (2.11)$$

where p is an integer (e.g., [149–151]). The above equation makes evident that $\partial_t \langle M_{\bullet} \rangle = 0$ and $\partial_t \langle M_{\bullet}^2 \rangle > 0$; in other words, the coagulation processes does not cause a drift in the mass distribution, but just enhances its dispersion by redistributing masses from the low to the high mass end. We mention that there is an exception to the mass conservation $\langle \dot{M}_{\bullet} \rangle = 0$ found above: when the kernel dependence on the mass is sufficiently strong, like in the well-known case of a multiplicative kernel $K(M_{\bullet}, M'_{\bullet}) \propto M_{\bullet} M'_{\bullet}$, then a sort of phase transition occurs, in that some mass may be lost in a finite time due to the appearance of clusters with infinite mass [158]. Such a phenomenon is called ‘gelation’ and has attracted a lot of interest in the scientific community for its practical applications to the physics of aerosol, polymers and molecular dynamics; however, this is not of our concern here (for a review, see [149]).

Note that the quantity $K(M_{\bullet,1}, M_{\bullet,2}) N(M_{\bullet,1}) N(M_{\bullet,2}) \equiv d^2 \mathcal{R} / dM_{\bullet,1} dM_{\bullet,2}$ just represents a merging rate density per bins of the two merging masses $M_{\bullet,1}$ and $M_{\bullet,2}$. However, in the present context it is more convenient to work with the merger rate density $d^2 \mathcal{R}_{\bullet \bullet \rightarrow \bullet} / dM_{\bullet,1+2} dq$ per bin of total (descendant) mass $M_{\bullet,1+2} \equiv M_{\bullet,1} + M_{\bullet,2}$ and mass ratio $q \equiv \min(M_{\bullet,1}; M_{\bullet,2}) / \max(M_{\bullet,1}; M_{\bullet,2})$, i.e. the ratio between the lesser and the more massive merging BHs (by definition). This can be achieved by a suitable change of variables in each term of the Smoluchowski equation, taking into account that the Jacobian of the transformation is just $M_{\bullet,1+2} / (1 + q)^2$. Identifying the two appropriate merging masses in each integral of equation (2.10), and splitting the destruction term into two parts to always have mass ratios $q < 1$, the overall Smoluchowski equation (2.10) can be rewritten in a

rather compact form as (see also [113])

$$[\partial_t N(M_\bullet, t)]_{\bullet\bullet\rightarrow\bullet} = \int_0^1 dq \frac{d^2 \mathcal{R}_{\bullet\bullet\rightarrow\bullet}}{dM_\bullet dq} \Big|_{M_\bullet} - \int_0^1 dq (1+q) \left[\frac{d^2 \mathcal{R}_{\bullet\bullet\rightarrow\bullet}}{dM_\bullet dq} \Big|_{M_\bullet(1+q)} + \right. \\ \left. + q \frac{d^2 \mathcal{R}_{\bullet\bullet\rightarrow\bullet}}{dM_\bullet dq} \Big|_{M_\bullet(1+q)/q} \right]. \quad (2.12)$$

Simulations indicate that the dependencies on mass, mass ratio and redshift of the merger rate densities per descendant dark matter halo or per descendant galaxy can be factorized [159, 160], so we make the simplifying assumption that the same holds for the hosted supermassive BHs. Therefore we parameterize the merger rate density per BH as

$$\frac{1}{N(M_\bullet, t)} \frac{d^2 \mathcal{R}_{\bullet\bullet\rightarrow\bullet}}{dM_\bullet dq}(M_\bullet, q, t) \simeq \mathcal{R}_{\bullet\bullet\rightarrow\bullet}(t) \left(\frac{M_\bullet}{10^{10} M_\odot} \right)^\eta p(q); \quad (2.13)$$

where η is a parameter specifying the mass dependence, dp/dq is the probability density distribution of mass ratios, and $\mathcal{R}_{\bullet\bullet\rightarrow\bullet}(t)$ is a redshift dependent normalization. Then the coagulation term turns into:

$$[\partial_t N(M_\bullet, t)]_{\bullet\bullet\rightarrow\bullet} = \mathcal{R}_{\bullet\bullet\rightarrow\bullet}(t) \left(\frac{M_\bullet}{10^{10} M_\odot} \right)^\eta \left\{ N(M_\bullet, t) - \int_0^1 dq p(q) (1+q)^{1+\eta} \left[N(M_\bullet, t) \Big|_{M_\bullet(1+q)} \right. \right. \\ \left. \left. + q^{1-\eta} N(M_\bullet, t) \Big|_{M_\bullet(1+q)/q} \right] \right\}. \quad (2.14)$$

We parameterize the redshift-dependent merger rate normalization in terms of a rescaling of the galaxy merger rate as estimated from galaxy pair counts observed by JWST (see [161–163]). Specifically, we take

$$\mathcal{R}_{\bullet\bullet\rightarrow\bullet}(t) = f_{\bullet\bullet\rightarrow\bullet} \mathcal{R}_0 (1+z)^\omega e^{-\tau(1+z)} \quad (2.15)$$

in terms of a scaling factor $f_{\bullet\bullet\rightarrow\bullet}$ and of the parameters $\mathcal{R}_0 \approx 0.013 \text{ Gyr}^{-1}$, $\omega \approx 3.4$ and $\tau \approx 0.14$ set by [162]. This choice is appropriate since it would be impossible to constrain the detailed redshift dependence of the BH merger rate relying only on the data concerning BH demographics and clustering, and on the mild integrated redshift dependence of the stochastic background in GWs (see below).

We anticipate that in our analysis $f_{\bullet\bullet\rightarrow\bullet}$ is not restricted to be less than unity, although it would be somewhat nonphysical to obtain a BH merger rate larger than the galactic one. However, this is done to account for observational scatter or possible systematics offsets in the galaxy merger rate. In fact, the latter is obtained combining the observed galaxy pairs fraction with a merger timescale extracted from numerical simulations, and thus inevitably subject to some uncertainties. Note that a different normalization of the merger rate will lead to a recalibration of $f_{\bullet\bullet\rightarrow\bullet}$ but will marginally affect the conclusions of the present work.

In addition, note that in the above we are neglecting any possible time delay between the merger of the BHs and that of the host galaxies, which constitutes the standard assumption in the analysis of the GW background [108, 164]; to wit, this maximizes the contribution of binary mergers.

Finally, following [112, 115] we take the mass ratio distribution $p(q) \propto q^{-\delta} \theta_H[q \geq q_{\min}]$ as a power-law with index δ extending down to a minimum mass ratio $q_{\min}$, held fixed to the reference values $\delta = -1$ and $q_{\min} = 0.1$. In fact, the dependence of the results on the mass ratio distribution is very weak, so there is no point in trying to fit the related parameters.

2.4 BH demographics

The direct output of the continuity equation is constituted by the BH demographics, i.e. the BH mass function at any cosmic time. This is found by solving the continuity equation:

$$\begin{aligned} \partial_t N(\log M_\bullet, t) = & -\frac{\partial_{\ln M_\bullet}}{t_{\text{Edd}}} \left[\int d\log \lambda p(\log \lambda) \lambda \frac{1-\epsilon-\epsilon_{\text{kin}}}{\epsilon} N_{\text{AGN}}(\log M_\bullet, t) \right] \\ & + \mathcal{R}_{\bullet \rightarrow \bullet}(t) \left(\frac{M_\bullet}{10^{10} M_\odot} \right)^\eta \left\{ N(\log M_\bullet, t) \right. \\ & \left. - \int_0^1 dq p(q) (1+q)^\eta \left[N(\log M_\bullet, t)_{|M_\bullet(1+q)} + q^{2-\eta} N(\log M_\bullet, t)_{|M_\bullet(1+q)/q} \right] \right\}. \end{aligned} \quad (2.16)$$

Note that although this is formally a partial integro-differential equation, the integrals only involve the mass function evaluated at the running time, so as to allow the exploitation of standard numerical techniques. In more detail, we solve equation (2.16) forward in cosmic time on discrete grids in redshift $z \in [0, z_{\text{ini}}]$ and BH masses $\log M_\bullet [M_\odot] \in [5, 11]$ with an explicit Runge-Kutta method of eight order.

Fiducially, we initialize the computation at $z_{\text{ini}} \sim 7$ which is the maximal redshift at which measurements of the AGN luminosity function by [79] do not require extrapolation, and there we assume as initial condition a mean duty cycle $\langle \delta \rangle(M_\bullet, z_{\text{ini}}) = 0.5$ independent of mass [29]. Then by definition of the duty cycle one has $N_{\text{AGN}}(\log M_\bullet, z_{\text{ini}}) = N(\log M_\bullet, z_{\text{ini}}) \langle \delta \rangle = N(\log M_\bullet, z_{\text{ini}})/2$ and hence, given the observed AGN luminosity function $N(L_{\text{AGN}}, z_{\text{ini}})$, one can determine the initial BH mass function $N(\log M_\bullet, z_{\text{ini}})$ just by solving equation (2.6). To explore the dependence of our results on the initial condition, we try to use $z_{\text{ini}} \sim 10$ by extrapolating the observed AGN luminosity functions. We also try to assume a different initial duty cycle in the range $\langle \delta \rangle(M_\bullet, z_{\text{ini}}) = 0.1 - 1$. Finally, we try a narrowly peaked mass function or the mass function stemming from various light-seed population models [165, 166]. Basically, we have found that as for the determination of the BH mass function at $z \lesssim 4$ and below, the initial redshift and reasonable initial conditions are marginally relevant since the accumulated mass from accretion soon largely exceeds that in the original BH population (see also [30, 132, 167]), and the impact of BH mergers on the latter at high redshift will turn out to be relatively minor.

2.5 BH clustering

A further observable to better constrain the parameters of our semi-empirical framework, and in particular to contribute in reducing the degeneracy between the parameters is to exploit the large-scale BH clustering [29, 61]. Specifically, we will be interested in comparing our model outcomes to data concerning the bias $b(M_\bullet, z)$, as provided at relatively low redshifts $z \lesssim 0.25$ by optical and X-ray observations of AGNs with robust BH mass determinations [168, 169].

In the context of our semi-empirical framework, the BH clustering signal can be computed by combining the BH mass function output from the continuity equation to the mass function and bias of dark matter halos as extracted from N -body simulations. In more detail, the procedure is the following. First of all, we employ the standard abundance matching technique [129, 170] to relate BH and halo masses in a nearly model-independent way, the only assumption being the existence of a monotonic mapping $M_H(M_\bullet, z)$ between the two variables, at least on statistical grounds. In particular, we use the version of abundance matching by [30] that includes the effect of the scatter in the halo-to-BH mass relation. This boils down to require the cumulative number densities in BHs and halos to match as:

$$\int_{\log M_\bullet}^{+\infty} d \log M'_\bullet N(\log M'_\bullet, z) = \int_{-\infty}^{+\infty} d \log M'_H N(\log M'_H, z) \times \frac{1}{2} \operatorname{erfc} \left[\frac{\log M_H - \log M'_H}{\sqrt{2} \sigma_{\log M_H}} \right], \quad (2.17)$$

and solve for $M_H(M_\bullet, z)$; here the scatter of the relation $\sigma_{\log M_H}$ dex is a fitting parameter. The integrand on the left hand side encases the BH mass function, that we obtain by solving the continuity equation. The right hand side involves the Λ CDM halo mass function $N(\log M_H, z)$, that we take from the N -body simulations by [171]. Note that this procedure is also supported by the recent pairwise residuals analysis by [172], which has highlighted that the correlation between BH and host halo mass appears as strong as the one with velocity dispersion, as corroborated by direct fitting (e.g., [173]) and clustering analysis (e.g., [174]). Finally, we employ the large-scale bias $b(M_H, z)$ of dark matter halos from the N -body simulations by [175], and convolve it with the halo mass vs. BH mass relation (properly taking into account its dispersion $\sigma_{\log M_H}$) to obtain the BH bias $b(M_\bullet, z)$ that can be then directly compared with clustering data. To speed up the above computation we actually exploit the accurate fitting formulas for the halo mass function⁷ and for the halo bias from the aforementioned N -body simulations, as implemented in the COLOSSUS package [176].

It is worth stressing that the procedure described above does not betray the data-driven spirit of our semi-empirical framework, because it does not require any major theoretical assumption apart from the underlying cosmological model (standard Λ CDM in our case) and the reasonable hypothesis that on statistical average larger BHs are associated to more massive halos. Note also the different way through which clustering is obtained within ab-initio models. These are also based on the dark matter halo catalogs (containing the halo clustering information) extracted from N -body simulations, but then the mapping between BH and halo masses requires to make substantial theoretical assumptions and to model the very complex baryonic physics involved in the formation and evolution of BHs. Such a challenging task is instead completely bypassed by our semi-empirical, data-driven framework based on the continuity equation.

2.6 Stochastic GW background

Binary mergers of supermassive BH throughout cosmic history can produce a stochastic background of GW with typical nano-Hz frequency, that is within the reach of present PTA

⁷We also include the correction for sub-halos in the halo mass function. However, this becomes relevant only at halo masses $M_H \lesssim 10^{12} M_\odot$ and thus it is found to play a minor role in the present context.

experiments. In the context of our semi-empirical framework, this can be easily computed from the outputs of the continuity equation, and in particular from the BH mass function and the BH merger rate. Specifically, the characteristic strain h_c of the stochastic GW background from binary BH mergers (in circular orbits) as a function of the observed frequency f is given by [177]

$$\begin{aligned} h_c^2(f) &= \frac{4G^{5/3}}{3\pi^{1/3}c^2f^{4/3}} \int \frac{dt}{(1+z)^{1/3}} \int dM_\bullet M_\bullet^{5/3} \int dq \frac{q}{(1+q)^2} \frac{d^2\mathcal{R}_{\bullet\bullet\rightarrow\bullet}}{dM_\bullet dq} = \\ &= \frac{4G^{5/3}}{3\pi^{1/3}c^2f^{4/3}} \int dt \frac{\mathcal{R}_{\bullet\bullet\rightarrow\bullet}(t)}{(1+z)^{1/3}} \int dM_\bullet M_\bullet^{5/3} \left(\frac{M_\bullet}{10^{10}M_\odot} \right)^\eta N(M_\bullet, t) \int dq p(q) \frac{q}{(1+q)^2}. \end{aligned} \quad (2.18)$$

where $N(M_\bullet, t)$ is the BH mass function obtained from solving the continuity equation (2.16) and $\mathcal{R}_{\bullet\bullet\rightarrow\bullet}$ is given by equation (2.15). Note that very often results of nano-Hz GW experiments are quoted in terms of a normalization $\mathcal{A}$ for the characteristic strain [178] written in the form $h_c(f) = \mathcal{A}(f/\text{yr}^{-1})^{-2/3}$, where the constant $\mathcal{A}$ can be computed easily from the expression above. The $f^{-2/3}$ scaling holds for circular orbits, and it can actually be altered by eccentricity or by environmental effects (in that they tend to somewhat reduce the strain and flatten its frequency dependence; see [179]), but these go beyond the scope of the present paper.

3 Data and analysis

We now turn to constrain the parameters of our semi-empirical framework by comparing its outcomes with various datasets. Specifically, we exploit local determinations of the BH mass function, of the BH clustering, and a recent estimate of the nano-Hz GW background.

As mentioned in section 1, the local BH mass function cannot be estimated directly, since the number of robustly measured BH masses is small (less than a hundred). Thus this is routinely obtained via the convolution of a galaxy statistics (measured on a much large sample of galaxies) with a scaling relationship between BH mass and galaxy properties (calibrated on the small sample of directly measured BH masses). Unfortunately, this procedure brings in a certain degree of uncertainty in the determination of the local BH mass function, which is important to take into account. To this purpose, we consider three cases. In the first case, the local BH mass function is obtained by combining the local stellar mass function or SMF $N(M_\star)$ by [180] with the $M_\bullet - M_\star$ relationship for early-type galaxies by [61, 181]. In the second case, the local BH mass function is obtained by combining the local galaxy velocity dispersion function or VDF $N(\sigma)$ by [182] and the observed $M_\bullet - \sigma$ relationship by [53]. Finally, in the third case the local BH mass function is obtained by combining the VDF with the ‘debiased’ version of the $M_\bullet - \sigma$ by [59]; this is meant to constitute a solid lower limit to the local demography, as pointed out by the recent results in [172]. Such binned local BH mass functions are reported in table 1. Our analysis will be conducted independently for each of these three cases, which confidently bracket the overall systematic uncertainty.⁸ They also span the range covered by other literature estimates from different galaxy statistics and/or

⁸Note that using the SMF and the intrinsic $M_\bullet - M_\star$ relationship by [59] produces a local BH mass function consistent with the combination VDF + intrinsic $M_\bullet - \sigma$, as shown explicitly by [61].

relationships. For example, the recent BH mass function by [183] strikes an intermediate course between our cases based on the observed $M_\bullet - \sigma$ and $M_\bullet - M_\star$ relations (see dotted line in figure 4). As another example, the determination by [79], obtained convolving the AGN luminosity function with the AGN luminosity vs. Eddington ratio relationship from SDSS [184, 185] and further correcting for obscured AGNs and inactive BHs, is in between our cases based on the observed and the debiased $M_\bullet - \sigma$ relations (see shaded area in figure 4).

As to BH clustering, we rely on the measurements of the large-scale bias $b(M_\bullet)$ as a function of BH mass at $z \sim 0.25$ by [168, 169] and at $z \sim 0.04$ by [186, 187]. Note that a wealth of additional datasets (e.g., [184, 186, 188–200]) are available at different redshifts for the clustering of AGNs as a function of luminosity from optical, X-ray and radio selected sample. However, we cannot rely on these datasets to independently test our framework. This is because we used as an input the AGN luminosity function, hence the relation between AGN luminosity and halo mass (obtained via abundance matching, see section 2.5) and the ensuing clustering signal as a function of AGN luminosity is reproduced by construction. Focusing instead on the clustering signal as a function of BH mass can independently constrain the halo to BH mass relationship and hence ultimately the BH mass function, which is the main output of our framework (see also [61, 187]).

As to the GW background, we rely on the most recent estimate from PTA by the **NANOGrav** collaboration [101], in terms of the strain amplitude parameter $\mathcal{A} = (2.4 \pm 0.7) \times 10^{-15}$ expressed by equation (2.18). We stress again that an added value of the present work is in attempting to fit *jointly* all these observables. In appendix B we will also test our model outcomes when excluding the PTA measurement from the likelihood, i.e. fitting only for BH demographics and clustering and then compare the resulting GW background with data.

We will also confront the outcomes of our semi-empirical framework with other auxiliary datasets not included in the fitting procedure: measurements of the local duty cycle $\langle \delta \rangle(M_\bullet)$ of AGN activity as a function of BH mass (taking into account appropriate luminosity threshold) by [135, 201, 202]; estimates of the active BH mass function at high-redshift $z \sim 4 - 6$, including BAL quasars and little red dots by [73, 93–96, 203, 204]; BH vs. halo mass relationship inferred from different methods [168, 205–207]; evolution with redshift of the average Eddington ratio for various kinds of AGNs [42, 44, 70, 89, 93–95, 185, 208–210]. Being affected by appreciable uncertainties and possibly by systematics or observational biases, these datasets are not employed to fit the parameters of our semi-empirical framework, but rather are exploited as *a-posteriori* validation.

For the analysis we adopt a Bayesian framework, characterized by the parameter set $\theta \equiv \{\epsilon_{\text{thin}}, \log \lambda_0, \xi, \eta, \log f_{\bullet \rightarrow \bullet}, \sigma_{\log M_H}\}$. We implement a Gaussian log-likelihood

$$\ln \mathcal{L}(\theta) = -\chi^2(\theta)/2, \quad (3.1)$$

where the chi-square $\chi^2(\theta) = \sum_{ij} [\mathcal{M}(M_{\bullet,i}, z_i|\theta) - \mathcal{D}(M_{\bullet,i}, z_i)] \mathcal{C}_{ij}^{-1} [\mathcal{M}(M_{\bullet,j}, z_j|\theta) - \mathcal{D}(M_{\bullet,j}, z_j)]$ is obtained by comparing our model expectations $\mathcal{M}(M_{\bullet,i}, z_i|\theta)$ to the data values $\mathcal{D}(M_{\bullet,i}, z_i)$, summing over different observables at their respective BH masses $M_{\bullet,i}$ and redshifts z_i and taking into account the variance-covariance matrix $\mathcal{C}_{ij}$ among different bins (here assumed diagonal). We adopt flat priors $\pi(\theta)$ on the parameters $\epsilon_{\text{thin}} \in (0.06, 0.42)$, $\log \lambda_0 \in [-3, 1]$, $\xi \in [0, 3]$, $\eta \in [-1, 1]$, $\sigma_{\log \lambda} \in (0, 1]$, $\log f_{\bullet \rightarrow \bullet} \in [-3, 1]$, $\sigma_{\log M_H} \in (0, 1]$. Note that $f_{\bullet \rightarrow \bullet}$ is

$\log M_{\bullet} [M_{\odot}]$	$\log N(\log M_{\bullet}) [\text{Mpc}^{-1} \text{ dex}^{-1}]$		
	$M_{\bullet} - M_{\star} [\text{obs.}]$	$M_{\bullet} - \sigma [\text{obs.}]$	$M_{\bullet} - \sigma [\text{deb.}]$
7.00	-2.71 ± 0.06	-2.59 ± 0.04	-2.70 ± 0.05
7.50	-2.60 ± 0.06	-2.69 ± 0.04	-2.88 ± 0.05
8.00	-2.62 ± 0.08	-2.87 ± 0.05	-3.18 ± 0.05
8.50	-2.79 ± 0.10	-3.17 ± 0.06	-3.65 ± 0.07
9.00	-3.18 ± 0.12	-3.63 ± 0.07	-4.34 ± 0.11
9.50	-3.83 ± 0.16	-4.32 ± 0.11	-5.30 ± 0.18
10.0	-4.76 ± 0.23	-5.28 ± 0.18	-6.59 ± 0.28

Table 1. The local BH mass functions exploited in our analysis. These include the determination from the stellar mass function (SMF) from [180] plus the $M_{\bullet} - M_{\star}$ relationship from [61, 181], the velocity dispersion function (VDF) from [182] plus the $M_{\bullet} - \sigma$ relationship in its observed [obs.] version from [53] or in its debiased [deb.] version from [59]. See text for more details.

not hard bounded by one, to take into account possible uncertainties in the galaxy merger rate. We will also consider a run where the supermassive BH merger rate follows unbiasedly the galaxy merger rate by setting a narrow gaussian prior on $\log f_{\bullet \rightarrow \bullet}$ centered in zero and with dispersion 0.1 dex. The overall set of parameters, the related equation in the main text where these first appear, and the adopted priors are reported in table 2.

We sample the parameter posterior distributions $\mathcal{P}(\theta) \propto \mathcal{L}(\theta) \pi(\theta)$ via the MCMC Python package `emcee` [211], running it with 10^5 steps and 100 walkers; each walker is initialized with a random position extracted from the priors discussed above. To speed up the computation, the active BH mass function is pre-determined solving equation (2.8) on a grid of redshift and Eddington ratio, and then its parameters are interpolated on the fly during the MCMC run. To more efficiently sample the parameter space, we adopt a mixture of differential evolution and snooker moves of the walkers, in proportion of 0.8 and 0.2 respectively, that emulates a parallel tempering algorithm. After checking the auto-correlation time, we remove the first 50% of the flattened chain to ensure burn-in.

3.1 Results and discussion

The results of our Bayesian analysis are displayed in figure 3 and are reported in table 3. Specifically, in figure 3 we illustrate the MCMC posterior distributions on our parameter set, for the different determinations of the local BH mass function (color-coded); the crosses highlight the positions of the bestfit and the marginalized distributions are normalized to unity at their maximum. In table 3 we summarize the marginalized posterior estimates of the parameters, in terms of their median values and 1σ credible intervals.

3.1.1 Parameter estimates

The radiative efficiency ϵ_{thin} is a well determined parameter, though it is very sensitive to the adopted local BH mass function determination: values $\epsilon_{\text{thin}} \approx 0.06$ close to the lower bound allowed for a thin disk are required when adopting the BH mass function based on the observed

Parameter	Prior	Equation
$\log \lambda_0$	$\mathcal{U}[-3, 1]$	(2.5)
ξ	$\mathcal{U}[0, 3]$	(2.5)
ϵ_{thin}	$\mathcal{U}(0.06, 0.42)$	(2.7)
η	$\mathcal{U}[-1, 1]$	(2.13)
$\log f_{\bullet\bullet\rightarrow\bullet}$	$\mathcal{U}[-3, 1]$ or $\mathcal{N}(0, 0.1)$	(2.15)
$\sigma_{\log M_{\text{H}}}$	$\mathcal{U}(0, 1]$	(2.17)

Table 2. Parameter set exploited in our Bayesian analysis. The first column lists the parameter symbol, the second column shows the related prior (the symbol $\mathcal{U}$ stands for a uniform distribution between the arguments, and $\mathcal{N}$ for a normal distribution with mean and dispersion given by the arguments), and the third column recall the equation of the main text where the parameter appears.

Parameter	$M_{\bullet} - M_{\star}$ [obs.]	$M_{\bullet} - \sigma$ [obs.]	$M_{\bullet} - \sigma$ [deb.]	$M_{\bullet} - M_{\star}$ [obs.] + $f_{\bullet\bullet\rightarrow\bullet} \approx 1$
ϵ_{thin}	$0.066^{+0.003}_{-0.008}$	$0.11^{+0.01}_{-0.01}$	$0.34^{+0.02}_{-0.03}$	$0.066^{+0.003}_{-0.008}$
$\log \lambda_0$	$-2.25^{+0.26}_{-0.23}$	$-1.43^{+0.22}_{-0.15}$	$-0.62^{+0.12}_{-0.08}$	$-2.22^{+0.27}_{-0.22}$
ξ	$2.57^{+0.64}_{-0.64}$	$1.03^{+0.38}_{-0.61}$	$0.36^{+0.06}_{-0.42}$	$2.53^{+0.58}_{-0.68}$
η	$0.59^{+0.44}_{-0.68}$	$0.64^{+0.48}_{-0.66}$	$0.65^{+0.58}_{-0.83}$	$0.52^{+0.33}_{-0.38}$
$\log f_{\bullet\bullet\rightarrow\bullet}$	$-0.77^{+0.8}_{-0.5}$	$-0.85^{+0.9}_{-0.6}$	$-1.4^{+1.1}_{-0.9}$	$-0.04^{+0.09}_{-0.09}$
$\sigma_{\log M_{\text{H}}}$	$0.68^{+0.04}_{-0.03}$	$0.58^{+0.04}_{-0.04}$	$0.25^{+0.08}_{-0.07}$	$0.68^{+0.04}_{-0.04}$
χ_r^2	1.22	1.24	1.25	1.47

Table 3. Results of our Bayesian analysis in terms of marginalized posterior estimates. For every parameter (listed in the first column) of our semi-empirical framework we report the median and the 2σ credible interval. The different columns refer to the different determinations of the local BH mass functions discussed in the main text. The last line lists the reduced χ_r^2 of the overall fit in the various cases.

$M_{\bullet} - M_{\star}$ relationship; values $\epsilon_{\text{thin}} \approx 0.1$ around the usually adopted one for the average population of AGNs (e.g., [64, 66]) are found when basing on the observed $M_{\bullet} - \sigma$ relationship; extreme values $\epsilon_{\text{thin}} \approx 0.3$ close to the upper bound allowed for a thin disk are required when adopting the BH mass function based on the debiased $M_{\bullet} - \sigma$ relationship. These variations are easily understood basing on the classic Soltan [62, 63] argument. As it is more easily seen in the simple case of equation (2.9), the local BH mass density $\rho_{\bullet} \equiv \int d \log M_{\bullet} M_{\bullet} N(\log M_{\bullet}, t)$ is essentially determined by the integral over cosmic time of the AGN luminosity density $\rho_{\bullet} \propto \frac{1-\epsilon-\epsilon_{\text{kin}}}{\epsilon} \int dt \int d \log L_{\text{AGN}} L_{\text{AGN}} N(\log L_{\text{AGN}}, t)$ scaled by the inverse of the radiative efficiency ϵ^{-1} . Thus BH mass function determinations featuring a higher normalization at the knee, which implies a larger BH mass density $\rho_{\bullet}$, require lower average efficiencies.

The values of the average efficiency estimated here could have far reaching physical consequences, since in the standard picture of disk accretion these can be traced back to the spinning state of the supermassive BHs [212, 213]. For example, if the local BH mass function from the debiased $M_{\bullet} - \sigma$ relationship would be the true one, the related high efficiencies $\epsilon \sim 0.3$ would correspondingly imply rapidly spinning behaviors in the typical supermassive

BH population. Unfortunately, this argument could not be considered as conclusive given the present uncertainties inherent to the local mass function determination.

The parameters λ_0 and ξ determining the local normalization and redshift evolution of the mean Eddington ratio are also rather well determined by our analysis. Here the relevant point is the mapping between the AGN luminosity function and the BH mass function via the Eddington ratio distribution that enters the continuity equation approach. Specifically, lower average values of λ tend to originate a flatter BH mass function at the high-mass end, since large BH masses correspond to moderate AGN luminosities (approximately $L_{\text{AGN}} \propto \lambda M_\bullet$ holds) falling in the flatter portion of the AGN luminosity function. Thus local BH mass function determinations with a flatter (steeper) high mass end prefer lower (higher) values of the average λ_0 . This explain the rising trend of the λ_0 estimate from the BH mass function determination based on the $M_\bullet - M_\star$ relationship, which is the flatter, toward that based on the debiased $M_\bullet - M_\star$ relationship, which features the steeper shape.

On the other hand, the parameter ξ ruling the redshift evolution of the mean Eddington ratio is strongly anti-correlated with λ_0 , so that lower values of λ_0 correspond to larger values of ξ and hence to stronger redshift evolution. This occurs because moving toward high redshift a progressively larger number of BHs are expected to be active (i.e., the duty cycle typically increases), so that the AGN luminosity function is more directly mapped into the relic BH mass function. In these conditions, values of λ substantially exceeding unity are not acceptable, since they would map high AGN luminosities and thus relatively small AGN number density to low BH masses; this will not allow an efficient buildup of the BH mass function at the low mass end.

The parameter η dictating the mass dependence of the binary merger rate of supermassive BHs is loosely constrained, though there is a tentative preference for positive values. This is mainly because of two reasons. On the one hand, as to the BH mass function, binary mergers tend to be appreciably relevant only at relatively low redshift $z \lesssim 1$ and at the very massive end $M_\bullet \gtrsim 10^9 M_\odot$; in fact, outside these ranges in redshift and BH mass, accretion is the main driver for the buildup of the mass function. On the other hand, as to the stochastic GW background, equation (2.18) shows that the parameter η appears in an integrated way as a correction to strong mass dependencies $M_\bullet^{5/3}$ associated to the characteristic GW strain. Therefore it is somewhat expected that the mass dependence in the merger rate can be only loosely constrained. However, it is somewhat interesting to note that bestfit values around $\eta \sim 0.5$ as found here would imply an increased merger rate for more massive BHs; this could reflect an intrinsic physical mechanism favoring them or a selection bias associated to the current nano-Hz GW background data.

The normalization scaling factor $f_{\bullet\bullet\rightarrow\bullet}$ of the supermassive BH merger rate with respect to the galaxy merger rate takes on median values $\lesssim 0.1$. These estimates are rather uncertain but a robust upper limit $\lesssim 1$ at $\gtrsim 2\sigma$ can be derived. The reason is the following. Given the observed bolometric AGN luminosity functions, accretion is able to reproduce by itself the BH demographics for reasonable values of the radiative efficiency and of the mean Eddington ratio. Thus the only solid constraint that our analysis can put on $f_{\bullet\bullet\rightarrow\bullet}$ is that it must not exceed appreciably unity; otherwise, given that the action of mergers is to redistribute masses

from the low to the high mass of the mass function, the latter will be flattened and the former lowered too much, in discrepancy with whatever local BH mass function determination.

Finally, the scatter in the $M_\bullet - M_H$ relationship inferred from abundance matching is robustly constrained in our analysis by the requirement to reproduce clustering measurements, i.e., the local bias as a function of BH mass. BH mass function determinations based on the observed $M_\bullet - M_\star$ or $M_\bullet - \sigma$ relationship call for considerable values of the scatter $\sigma_{\log M_H} \gtrsim 0.5$ dex between BH and halo mass; on the other hand, the BH mass function determination based on the debiased $M_\bullet - \sigma$ relation requires appreciably lower values $\sigma_{\log M_H} \gtrsim 0.3$ dex. This is because scaling relations of BH to galaxy properties with lower normalization tend to map on the average a given BH mass into a larger halo mass, hence would imply a higher bias; as a consequence, the scatter in BH to halo mass cannot be too large otherwise the predicted clustering signal would exceed the observational constraints.

3.1.2 BH demographics and clustering

In figure 4 we illustrate the BH mass function resulting from our semi-empirical framework. Solid colored lines show the bestfits to the three local BH mass function determinations (colored symbols) based on the observed $M_\bullet - M_\star$ (red), observed $M_\bullet - \sigma$ (blue) and debiased $M_\bullet - \sigma$ (green) relationship. In addition, the magenta line is the bestfit to the $M_\bullet - M_\star$ determination when a narrow prior on the scaling parameter of the BH to galaxy merger rate $f_{\bullet\rightarrow\bullet}$ around unity is adopted. All the fits are remarkably good, highlighting that the parameterizations adopted in our semi-empirical framework are sensible and sufficient to describe the BH demographics. Moreover, the dashed colored lines illustrate the corresponding active BH mass function at high redshift $z \sim 5$, which should be compared with the observational estimates by [73, 93–96, 203] at $z \sim 4 - 6$ (dark grey symbols). Note that the high-redshift active BH mass function has not been included in the fitting process, so the good agreement found with these independent data is a genuine prediction of our model, and further corroborates our choice of a redshift dependent mean Eddington ratio.

In figure 5 we compare the large-scale bias $b(M_\bullet, z)$ as a function of BH mass from our semi-empirical framework for the different local BH mass function determinations (colored lines) to clustering data at $z \sim 0.01 - 0.25$ by [168, 169, 186, 187], finding an excellent agreement. In addition, the inset shows the BH bias as a function of redshift, after averaging over the supermassive BH mass function, i.e. computing $\langle b \rangle(z) \equiv \int d \log M_\bullet N(\log M_\bullet, z) b(M_\bullet, z) / \int d \log M_\bullet N(\log M_\bullet, z)$. Although this redshift dependence of the bias has not been fitted upon, the outcomes from our semi-empirical framework agree reasonably well with the observational estimates from optical (plus signs; [184, 189, 197]), X-ray (diamonds; [188, 191, 192]) and radio-selected (inverse triangles; [194, 199, 200]) samples.

In figure 6 the related local BH to halo mass relationship is displayed, and compared with various observational estimates by [168, 205–207] derived by a variety of methods (e.g., clustering, kinematics, weak lensing). Actually the data are rather dispersed, partly because of intrinsic variance in the relationships, and partly because of possible observational systematics. In particular, there is a tendency for weak lensing estimates to predict a BH to halo mass relation with a lower normalization with respect to clustering measurements. Note that in view of these observational uncertainties, the BH to halo mass relation has not been

fitted upon in our analysis. However, the relation derived via abundance matching from our semi-empirical framework, also in view of the estimated intrinsic scatter $\sigma_{\log M_{\text{H}}} \gtrsim 0.3 - 0.5$ dex (colored errorbars in the plot), is consistent with the cloud of datapoints.

Figure 7 illustrates the local BH duty cycle as a function of BH mass from our semi-empirical framework for the different local BH mass function determinations (colored lines), compared with classic observational estimates by [135, 201, 202] derived from AGN active fractions in the local Universe. First of all, it is important to notice that there is a rather large dispersion in the data, which partly reflects differences in bolometric corrections, luminosity thresholds, and systematic offsets in the BH mass estimates; for these reasons the duty cycle has not been fitted upon in our analysis. Nevertheless, the duty cycle reconstructed from the continuity equation is found to be broadly consistent with these observational estimates. More in detail, higher duty cycle around $1 - 10\%$ are produced for the models fitting the local BH mass function determination based on the observed $M_{\bullet} - M_{\star}$ relation, while smaller duty cycles $\sim 0.1 - 1\%$ are originated in the model fitting BH mass function based on the debiased $M_{\bullet} - \sigma$ relationship.

In the top panel of figure 8 we showcase the mean Eddington ratio as a function of redshift as derived from our analysis for the different local BH mass function determinations (colored lines). These outcomes are compared with a multitude of observational estimates from X-ray/mid-IR/optical AGNs by [70, 81, 185, 208], broad-absorption-line quasars by [42, 44, 94, 95], and little red dots by [89, 93, 209, 210]. The various datasets are somewhat heterogeneous, so that the comparison with the model outcomes should be taken only as indicative. In fact, we stress that the parameters λ_0 and ξ determining in our model the redshift-dependent Eddington ratio after equation (2.5) have not been fitted on the data reported here, but instead have been determined by reproducing the BH demographics and clustering via the continuity equation approach. Nevertheless, it is reassuring that the model outputs on $\lambda(z)$ are broadly consistent with these observational determinations for individual AGNs. As already mentioned, the outcome referring to the BH mass function determination based on the observed $M_{\bullet} - M_{\star}$ relation tends to predict rather low mean Eddington ratios $\lambda \lesssim 0.01$ in the local Universe and a steep rise toward high redshift values $\lambda \sim 1$ at $z \sim 6$. This behavior agrees well with the average Eddington ratio observationally determined for type-1 AGNs up to $z \lesssim 4$ by [72], as fitted by [31]; it also accords with the dependence assumed on phenomenological grounds by [29]. On the other hand, the model outcomes referring to the BH mass function determinations based on the observed or debiased $M_{\bullet} - \sigma$ relations point toward somewhat higher current normalization $\lambda \lesssim 0.3$ and an appreciably weaker evolution toward high redshift. This behavior reflects the well-known degeneracy between the local duty cycle and the evolution of the Eddington ratio (see [29]). In the bottom panel of figure 8 we illustrate the full Eddington ratio distribution at three representative redshifts $z \approx 0.8$ (left) 1.8 (middle) and 3.2 (right). The outcomes of our semi-empirical framework are in good accord with the estimates for type-1 AGNs by [72] (circles) and with the corresponding lognormal fits by [31] (dashed lines).

In figure 9 we illustrate the relative contributions of mergers to the increase in the BH mass function, sliced in different bins of BH mass and redshift. The color scale illustrates the ratio between the production rate of the mass function due to mergers (i.e., the positive term

in the Smoluchowski equation) normalized to the sum of this quantity and of the analogous one due to accretion. It is evident that the main impact of mergers in increasing the mass function occurs toward low redshift $z \lesssim 1.5$ and for relatively high masses $M_\bullet \gtrsim \text{some } 10^8 M_\odot$. Such a contribution is extended somewhat toward higher redshift and smaller BH masses for the BH mass function determination based on the $M_\bullet - M_\star$ relation, which features the highest normalization at the high mass end, and thus leave more rooms for redistribution of the masses there. An additional yet modest contribution occurs at intermediate redshift $z \sim 2 - 4$ because the redshift increase in the BH (and galaxy) merger rate cooperates with a still appreciably high BH mass function; at higher z the effect is less evident since the merger rate stays put but the BH mass function drops considerably. Finally, it is worth noticing that for very large BH masses $M_\bullet \gtrsim 10^9 M_\odot$ mergers are quite relevant even at substantial redshifts $z \gtrsim 6$, though clearly such over-massive BHs in the primordial Universe would be rare.

3.1.3 Stochastic GW background

In figure 10 we illustrate our posterior estimates for the normalized stretch amplitude of the nano-Hz GW background for the different local BH mass function determinations (colored violin plots), with the black dots and solid errorbars illustrating the median and the 2σ credible intervals. These should be compared with the observational estimate of the background amplitude by the **NANOGrav** collaboration [101] (yellow solid line with shaded area). It appears evident that the outcomes of our analysis do not favor a substantial contribution to the observed GW background from supermassive BH mergers. Even in the extreme case (magenta violin plot) where we have adopted the BH determination based on the $M_\bullet - M_\star$ relation (which has the highest normalization) and we have imposed a narrow prior around unity on the rescaling parameter $f_{\bullet\rightarrow\bullet}$ between the BH and the galaxy merger rate, our analysis predicts that only a fraction $\lesssim 30 - 50\%$ (at $\sim 3\sigma$) of the observed background can be traced back to supermassive BH mergers. Our results are in good agreement with the outcomes of the **Astrid** and **TNG-300** numerical simulations (see [117]) and of the **holodeck** suite (see [101]); in fact, the latter two have a median closer to our maximized result with $f_{\bullet\rightarrow\bullet} \approx 1$, but the large uncertainties make them consistent also with our more conservative estimates.

In figure 11 we highlight the contribution of supermassive BH mergers to the nano-Hz GW stochastic background sliced in redshift and BH mass, for the different local BH mass function determinations (different panels). Most of background is contributed by relatively low redshift $z \lesssim 3$ sources and large BH masses $M_\bullet \gtrsim 10^9 M_\odot$. Such a contribution extends toward higher redshift and smaller BH masses for the BH mass function determination based on the $M_\bullet - M_\star$ relation, which features the highest normalization at the high mass end; this behavior is then reinforced when the binary BH merger rate is maximized and imposed to follow the galaxy merger rate (i.e., when $f_{\bullet\rightarrow\bullet} \approx 1$; top right panel). This figure also highlights that the tension between BH demographics and GW background is robust to residual uncertainties on the basic inputs of our data-driven model. For example, one could implement variations in the AGN luminosity function at $z \gtrsim 4.5$ (see appendix A), which may account for a steeper trend/higher normalization with respect to [79] as claimed by [214–217] or possibly suggested by an increased UV space density due to little red dots in case these are truly AGN dominated (see [96]). However, the impact on the GW background from

such changes at relatively high redshift and small to moderate AGN luminosities/BH masses would be minor (more quantitative analysis is presented in appendix B).

A possible tension between the GW background expected from BH mergers and the one measured by PTA has been noted and discussed before in the literature (e.g., [101, 108, 110, 113, 117, 164, 218–221]). However, these studies lack a self-consistent treatment of BH mergers, in that the related rates (usually derived from galaxy merger rates via a BH-galaxy scaling relation) are used to compute the expected GW background, but are not employed at the same time to build up and check the related impact on the mass function. Our analysis point out the tension via a semi-empirical, data-driven model based on the continuity plus Smoluchowski equation, that treats unitarily accretion and mergers, incorporates the latest observational determinations of the AGN luminosity functions and galaxy merger rates, and is jointly constrained by local BH demographics and GW background.

The situation is better clarified in figure 12, where we illustrate the merger rate per descendant BH (basically equation (2.13) integrated over the mass ratio distribution) vs. redshift, referred to a reference BH mass $M_{\bullet} \sim 10^{10} M_{\odot}$ which is taken as a representative contributor to the GW background. Note that the mass dependence is anyway not an issue since it is weak and the range of BH masses contributing to the background is rather limited (see above). Colored solid lines illustrate the median outcomes from our semi-empirical framework and MCMC analysis, for the different local BH mass function determinations. Essentially they are rescaled version (by the parameter $f_{\bullet \rightarrow \bullet}$) of the solid gray curve, which represents the galaxy merging rate as estimated from pairs counts by JWST. However, note that the 2σ uncertainty (not reported for the sake of clarity otherwise the plot would look crowded) for the red, blue and green cases essentially extends from the magenta line downwards, since our model can only set an upper limit to the merging rates per BH, as it is evident from the posterior of the parameter $f_{\bullet \rightarrow \bullet}$ in figure 3. For comparison, the dashed gray line refers to the major merger rate per galaxy from the *Illustris* simulation by [222], converted to the reference BH mass via the BH-stellar mass relation by [52], while the dotted gray line is the major merger rate per dark matter halo by [223], converted to the reference BH mass via the BH-halo mass relation by [173].

The take-home message is that to reproduce the high mass end of the estimated local mass functions (regardless of the specific determination), our model restricts the merger rate to be smaller or in any case not to exceed that inferred from galaxy pairs by [162, 163]. This is in turn slightly lower at $z \lesssim 0.5$ than that measured for galaxies in hydrodynamic simulations and for halos in N -body codes. Contrariwise, fitting the GW background measured by PTA experiments would require the merger rate per BH to be appreciably higher, as illustrated by the yellow curves that refer to various parametric analyses of the PTA data by [164, 218, 220]. However, these high rates would inevitably corrupt the local BH mass function at the high mass end, in a way that cannot be effectively cured by any reasonable accretion-related parameter set. Moreover, these merger rates per BH required to fit the PTA data turn out to be appreciably higher than the merger rate per galaxies or per halo expected from observations of galaxy pairs and from numerical simulations, an occurrence which would be in itself rather problematic to explain.

In principle, one could envisage physically-motivated modifications of the current setup to potentially ease the tension between the PTA signal and the model predictions. For example, a local BH mass function more skewed toward larger masses than the ones presented in this work could provide room for more frequent and/or more massive BH mergers, thus boosting the implied GW background signal. This could originate from a particularly flat stellar mass function at the high-mass end, and/or from an appreciably high normalization of the $M_\bullet - \sigma$ relationship (as may be the case for cored ellipticals, see [224]; also K. M. Varadarajan in preparation). Another possibility may involve a radiative efficiency increasing with the BH mass (see [29, 225]) which progressively reduces the number density of the most massive BHs, so as to reconcile the local BH mass function with the one derived from accretion even in the event of an increased number of BH mergers. However, a couple of reasons suggest that these attempts are likely doomed to fail. First, the merger rate density of galaxies is nowadays well constrained by JWST data, and even requiring the merger rate density of BHs to mirror it is not sufficient to reproduce the PTA background, as shown by our analysis. Therefore, independently of the shape of the local BH mass function, explaining all the GW background via BH mergers would require to assume that the merger rate per BH is appreciably larger than that of galaxies. Second, leaving more room to mergers at late times in building up the local BH mass function at the high-mass end could be dangerous, since mergers would also tend to reduce the number of low-mass BH (which merge into larger ones) so distorting the mass function there. All in all, solving the tension by acting solely on the accretion-related parameters would require some fine-tuning and would remain a very challenging task. We plan to perform an in-depth analysis of these and further possibilities in a dedicated future work.

4 Summary

We have explored the evolution of the supermassive BH population via a semi-empirical approach with minimal assumptions and data-driven inputs. This is based on a continuity plus Smoluchowski equation framework aimed to describe jointly the two main modes of BH growth, namely gas accretion and binary mergers. We have educatedly parameterized the main uncertainties associated to each BH growth channel, and inferred constraints on these in a Bayesian setup from estimates of the local BH mass function, clustering, and measurements of the nano-Hz stochastic GW background. Our main findings are as follows.

- Different local mass function determinations imply substantially different values for the parameters regulating the accretion mode of the supermassive BH population. The BH mass function based on the stellar mass function and observed $M_\bullet - M_\star$ relationship requires low radiative efficiency $\epsilon \sim 0.06$, small local average values of the Eddington ratio $\lambda \sim 10^{-2}$ with a rather strong dependence on redshift $\lambda \propto (1+z)^{2.5}$; a high dispersion $\sigma_{\log M_H} \sim 0.7$ in the relationship between BH and host halo mass is also required to reproduce the BH clustering. The BH mass function based on the velocity dispersion function and the observed $M_\bullet - \sigma$ relationship allows for standard values of the efficiency $\epsilon \sim 0.1$, local average Eddington ratios $\lambda \lesssim 10^{-1}$ with a mild redshift evolution $\lambda \propto (1+z)$, and scatter of $\sigma_{\log M_H} \sim 0.6$ in the BH-halo mass relation. Finally, the BH mass determination based on the debiased $M_\bullet - \sigma$ relationship

requires considerably high values of the efficiency $\epsilon \sim 0.3$, local average Eddington ratio $\lambda \lesssim 0.3$ modestly evolving with redshift $\lambda \propto (1+z)^{0.4}$, and appreciably lower scatter $\sigma_{\log M_H} \sim 0.3$ in the BH vs. halo mass relation. The difference in efficiency could have far-reaching implications on the BH spinning state of the BH population, though systematic uncertainties in the local BH mass function determination do not allow to draw firm conclusions yet.

- In all the above instances the fits from our semi-empirical framework to the local BH mass function and BH clustering are remarkably good. Moreover, the outcomes of our analysis are able to reproduce auxiliary observables that were not included in the fitting procedure: the active BH mass function at high redshift $z \sim 3-6$ comprising the recently discovered population of little red dots, the local BH vs. halo mass relationship, the local BH duty cycle and the evolution of the mean Eddington ratio across cosmic times. Such an agreement supports our semi-empirical, data-driven framework and the outputs of the continuity equation approach.
- Regardless of the adopted local BH mass function determination, we have found that the impact of binary BH mergers on the BH demographics is rather limited, and becomes only relevant for BH masses $M_\bullet \gtrsim \text{some } 10^8 M_\odot$ at relatively low redshifts $z \lesssim 1.5$. This result stems from the fact that most of the local BH mass density is acquired via gas accretion, with mergers playing a secondary role at low redshift and large BH masses. More quantitatively, the best-fit values for the rescaling factor of the BH to galaxy merger rate is $f_{\bullet \rightarrow \bullet} \lesssim 10^{-1}$. These estimates are quite uncertain but a robust upper limit $\lesssim 1$ can be derived at $\gtrsim 2\sigma$. There is a tentative evidence for a weak direct dependence of the binary BH merger rate on mass $M_\bullet^{\sim 1/2}$, though the uncertainty on the slope is still very large.
- Our analysis does not favor a substantial contribution from supermassive BH mergers to the observed GW background. Irrespective of the local BH mass function determination, we have constrained the contribution of binary BH mergers to be less than $\lesssim 30\%$ (at $\sim 3\sigma$) of the measured background. Even in the extreme case where we adopt the BH determination with the highest normalization at the massive end (the one based on the $M_\bullet - M_\star$ relation) and a BH merger rate strictly mirroring the galactic one (i.e. $f_{\bullet \rightarrow \bullet} \approx 1$), we have bound the contribution from BH mergers to the GW background at around $\lesssim 50\%$ at $\sim 3\sigma$. Finally, we have highlighted that binary BH mergers at relatively low redshift $z \lesssim 3$ with large BH masses $M_\bullet \gtrsim 10^9 M_\odot$ are the main contributors to the observed GW background.

In a future perspective, it would be beneficial to input our semi-empirical framework with improved determinations of the AGN luminosity functions and active BH mass functions, especially toward high redshift. In this respects, new data from *Euclid*, *JWST*, *e-ROSITA* and eventually *NewATHENA* will clear residual uncertainties on bolometric corrections, obscured AGN fraction, and abundance of BHs with faint bolometric luminosities. In addition, it would be extremely important to solve the longstanding debate concerning the local BH mass function determination. Although to build a BH mass function based on an extended

sample of quiescent BHs with robustly determined masses would remain challenging in the near future, a great step forward could be to identify the most fundamental BH-galaxy relationship and assess the observational and systematics issues that may affect and bias it. Wide galaxy surveys that will be conducted by the **Rubin-LSST** and eventually by **ELT** will help in this respect. All the above will improve appreciably the constraints on the parameters regulating the accretion mode for the growth of the BH population, like the AGN luminosity function, the radiative efficiency (hence BH spins), and the redshift-dependent Eddington ratio distributions.

As for the impact of binary BH mergers on the growth of the supermassive BH population, in the future it will be effectively constrained by more precise measurements of the nano-Hz GW background from PTA experiments; this will be possible thanks to the advent of the **SKAO** and its precursors, that will allow the monitoring of a substantially increased amounts of pulsars. These data will greatly improve the constraints on the mass and possibly the redshift dependence of the BH merger rate, allowing to confirm the conclusion of the present analysis concerning a modest contribution of binary BH mergers to the GW background. If so, the implications would be enormous. On the one hand, this would be a good news for **LISA**, since supermassive BH mergers may originate an appreciable background noise in the mHz band, degrading the effective sensitivity of the antennae to many primary science goals of the mission (e.g., [226]), like the detection of individual massive BH binaries, verified galactic binaries, and extreme mass ratio inspirals. On the other hand, our current results could motivate an exciting search for alternative astrophysical sources or even rise the hype for a primordial, cosmological origin of the remaining signal measured by PTA experiments.

Acknowledgments

We thank the anonymous referee for constructive and helpful comments. This work was partially funded from the projects: INAF GO-GTO Normal 2023 funding scheme with the project “Serendipitous H-ATLAS-fields Observations of Radio Extragalactic Sources (SHORES)”; INAF Large Grant 2022 project “MeerKAT and LOFAR Team up: a Unique Radio Window on Galaxy/AGN co-Evolution”; INAF Large GO 2024 project “MeerKAT and Euclid Team up: Exploring the galaxy-halo connection at cosmic noon”; “Data Science methods for MultiMessenger Astrophysics & Multi-Survey Cosmology” funded by the Italian Ministry of University and Research, Programmazione triennale 2021/2023 (DM n.2503 dd. 9 December 2019), Programma Congiunto Scuole; Italian Research Center on High Performance Computing Big Data and Quantum Computing (ICSC), project funded by European Union — NextGenerationEU — and National Recovery and Resilience Plan (NRRP) — Mission 4 Component 2 within the activities of Spoke 3 (Astrophysics and Cosmos Observations); European Union — NextGenerationEU under the PRIN MUR 2022 project n. 20224JR28W “Charting unexplored avenues in Dark Matter”. MB acknowledges that this article was produced while attending the PhD program in PhD in Space Science and Technology at the University of Trento, Cycle XXXIX, with the support of a scholarship financed by the Ministerial Decree no. 118 of 2nd March 2023, based on the NRRP — funded by the European Union — NextGenerationEU — Mission 4 “Education and Research”, Component 1 “Enhancement of the offer of educational services: from nurseries to universities” — Investment

4.1 “Extension of the number of research doctorates and innovative doctorates for public administration and cultural heritage” — CUP E66E23000110001. LB acknowledges financial support from the German Excellence Strategy via the Heidelberg Cluster of Excellence (EXC 2181 — 390900948) STRUCTURES. HF acknowledges support at Fudan University from the Shanghai Super Post-doctoral Excellence Program grant No. 2024008. DR acknowledges support from the University of Southampton via the Mayflower studentship.

A The high-redshift AGN luminosity function

In the main text we have adopted as our fiducial input the AGN luminosity function by [79], and in particular their fitting formula reported in equation (2.2). This currently constitutes the best homogenized rendition of different multi-wavelength datasets on the AGN luminosity function, converted to bolometric and accounting for obscured fraction with common corrections. However, recently the shape of the [79] luminosity function at $z \gtrsim 4.5$ has been somewhat challenged by new estimates from optical and X-ray data (see [209, 210, 216, 217, 227]). The issue has been also reinforced by the population of overmassive, X-ray underluminous BHs recently discovered by JWST (e.g., [87–91, 228]), partly located in the so called ‘little red dots’ sources (see [92–97]).

The current situation may be visualized in figure 13, where the AGN luminosity function by [79] at redshift $z \sim 5 - 7$ (color-coded) is illustrated and compared to the most recent data collected by [217]. Besides the still large uncertainties in the data, there is a tendency for the fit by [79] to somewhat under-predict the observational estimates. Remarkably, we checked that the rendition by [79] can be reconciled with these new data just by modifying the parameter controlling the number density evolution $d_1 \rightarrow d_1 + 0.2$ for $z \gtrsim 4.5$ (i.e., the negative number density evolution is slowed down). The results are illustrated by the solid lines, which agree with the data remarkably better.

One may wonder what would be the impact of this revised fit to the AGN luminosity function on the results presented in the main text. We have checked that this is irrelevant for the building up of the local BH mass function, since most of the accreted mass is acquired at redshift $z \lesssim 4$. In addition, the correction also turns out to be negligibly important for the GW background in the nano-Hz band, since most of it is originated at $z \lesssim 3$ (see figure 11). A more quantitative analysis is presented in appendix B.

B Additional tests

In this appendix we present additional tests on the robustness of the result on the expected GW background to variations in some assumptions of our analysis. We rely on the local BH mass function determination based on the stellar mass function and the $M_\bullet - M_\star$ relation, since this tends to maximize the impact of binary BH mergers. We focus on how the posterior on the GW normalized stretch amplitude is altered when the following assumptions are modified with respect to our fiducial analysis in the main text.

- AGN luminosity functions at high redshift. In the main text we have used the analytical rendition by [79] over the redshift range $z \sim 0 - 7$. Here we try to employ the correction

to the $z \gtrsim 4.5$ behavior discussed in appendix A which is suggested by the most recent high- z data.

- Kinetic efficiency ϵ_{kin} . In the main text we have assumed $\epsilon_{\text{kin}} = 0.15 \epsilon_{\text{thin}}$ following the analysis by [144]; given the bestfit values of $\epsilon_{\text{thin}} \lesssim 0.1$, rather small kinetic efficiencies are implied. However, since it has been claimed (e.g., [145, 146]) that the kinetic efficiency could be larger, here we test the impact of assuming $\epsilon_{\text{kin}} = 0.2$.
- PTA measurement in the likelihood. In the main text we have included the GW background signal by **NANOGrav** as a proper measurement in the likelihood. Here we remove it from the likelihood, so that we just fit for the BH demography and clustering, and then check the GW background *a-posteriori*.

The outcomes of these variations are illustrated in figure 14. Plainly, no major alteration with respect to the fiducial result of the main text is originated. In particular, adopting the corrected version of the high- z luminosity function (which has an higher normalization) implies more mass added by accretion in the early buildup of the BH mass function, hence somewhat less room for mergers toward low redshifts where they become important. On the other hand, raising the kinetic efficiencies to values $\epsilon_{\text{kin}} = 0.2$ lowers somewhat the mass accumulated by accretion, and leave more room for mergers; however, the effect is small since the relevant combination of parameter $(1 - \epsilon - \epsilon_{\text{kin}})/\epsilon$ depends mostly on the radiative efficiency ϵ than on the kinetic ϵ_{kin} . In addition, the change may be more effectively (in terms of the fit on the local BH mass function and clustering) compensated by varying the Eddington ratio distribution parameters than by raising the merger rates. Finally, eliminating the PTA measurements from the likelihood reduces the predicted background because the local BH mass function is already fitted well with the mass accumulated by accretion; thus the number of mergers is kept to values smaller than in the fiducial case, where instead the PTA measurement in the likelihood allows it to higher values, though still compatible with the BH demographics.

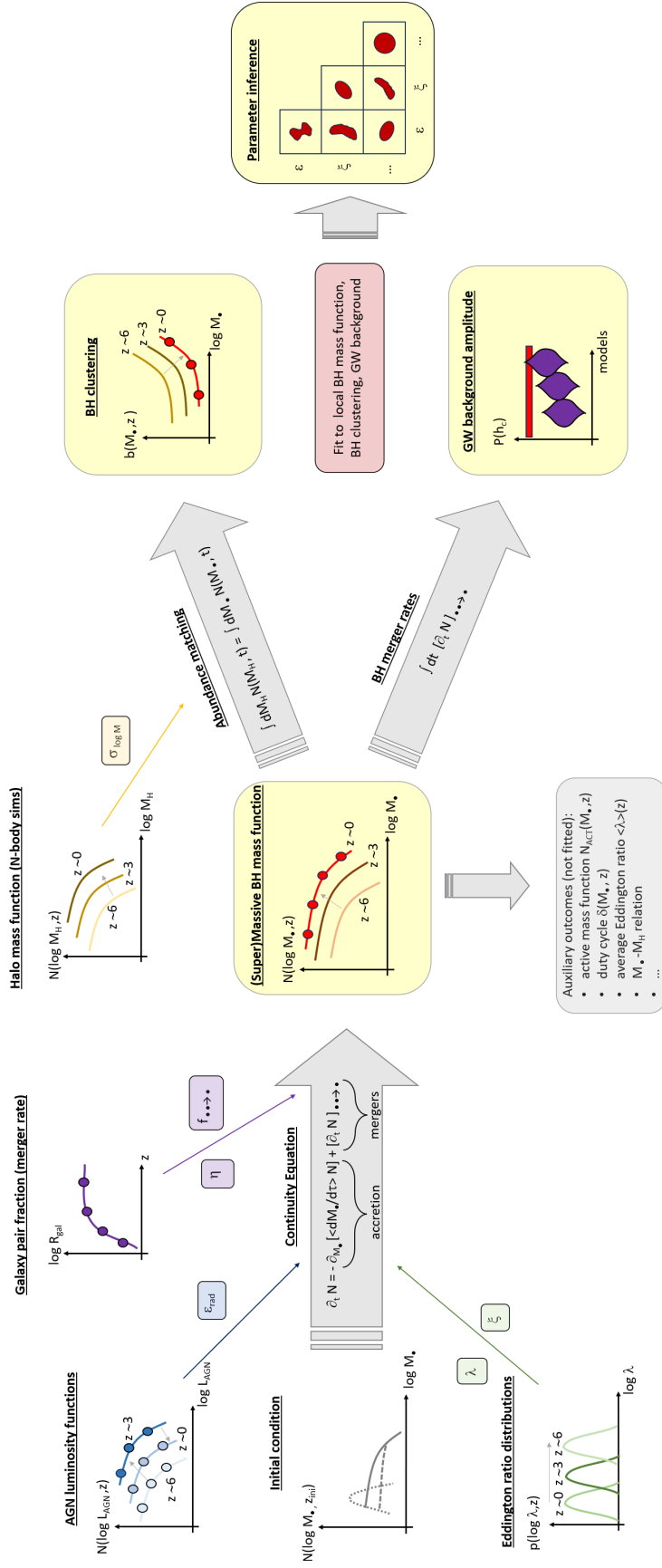


Figure 1. Schematics showing the basic methodology underlying our semi-empirical framework (see section 2 for details).

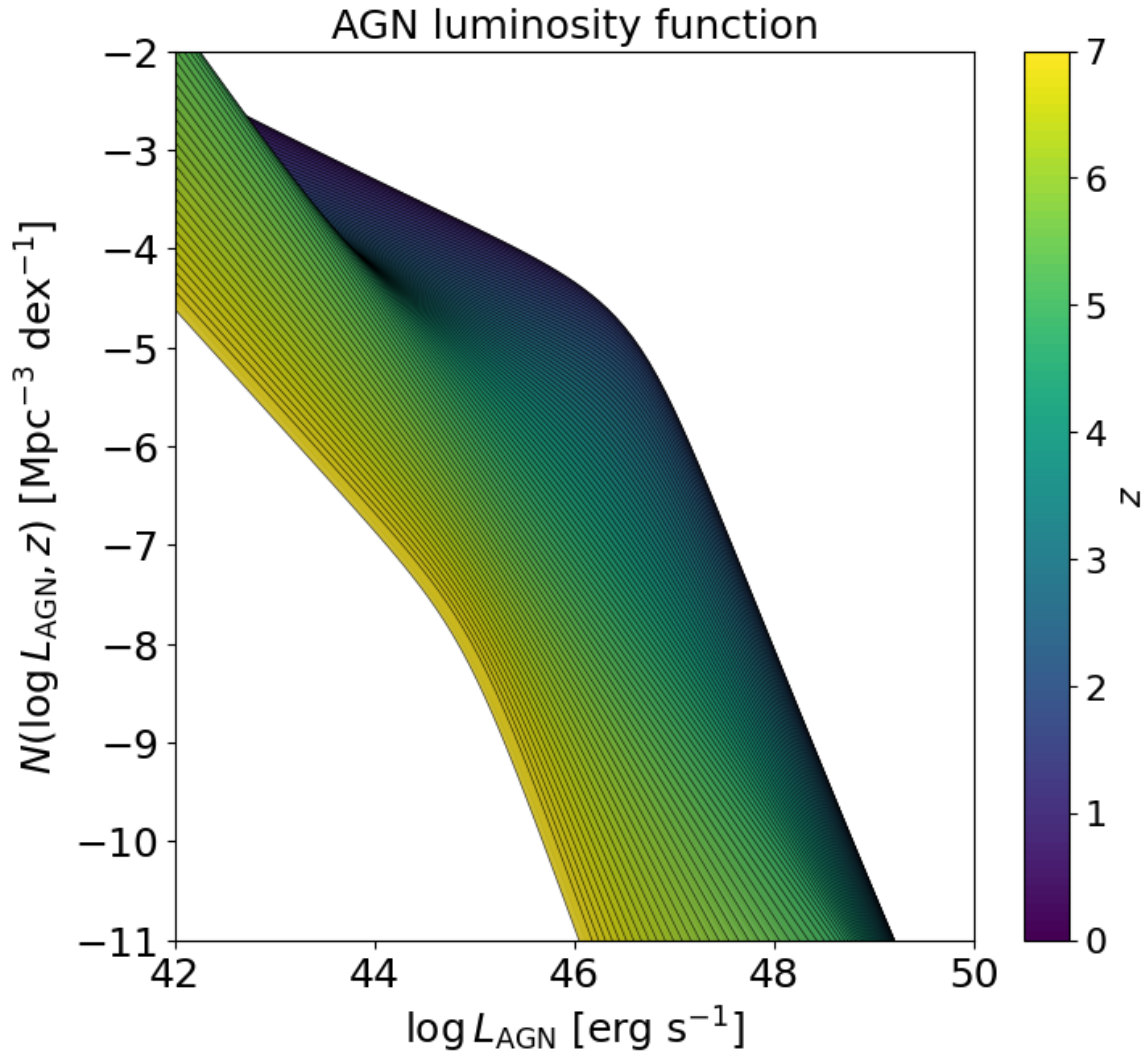


Figure 2. The AGN bolometric luminosity functions derived from multi-wavelength data by [79] (we use the fit labeled ‘A’ in their table 4; see also section 2.2) as a function of redshift (color-coded), which is the basic input in our semi-empirical framework.

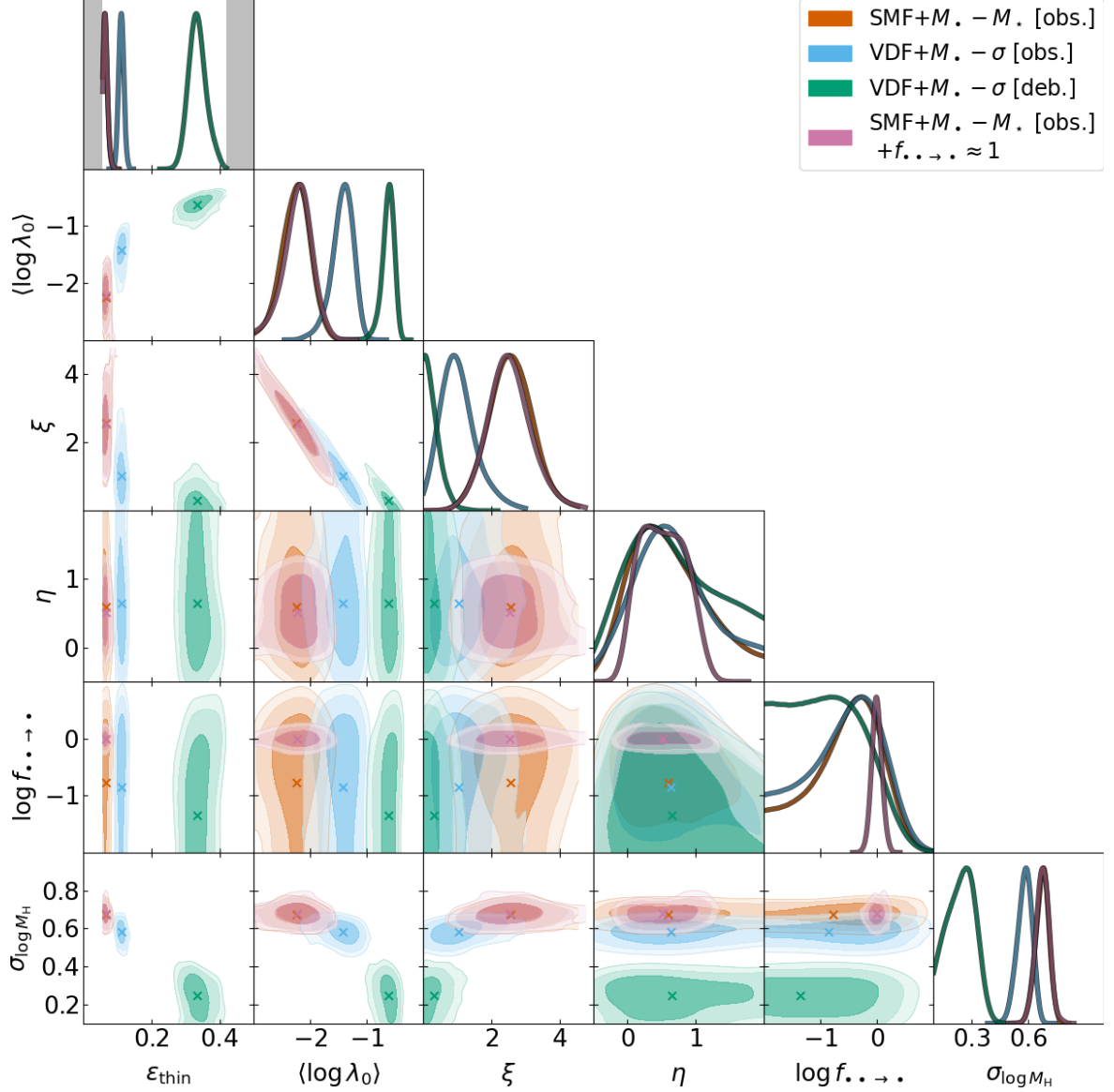


Figure 3. MCMC posterior distributions for the parameters exploited in our semi-empirical framework: radiative efficiency of the thin disk phase ϵ_{thin} ; mean Eddington ratio $\langle \log \lambda_0 \rangle$ and its redshift evolution parameter ξ ; mass-dependence parameter η in the supermassive BH merger rate and rescaling parameter $f_{\bullet \rightarrow \bullet}$ of the BH to galaxy merger rate; scatter $\sigma_{\log M_H}$ in the abundance matching relationship between halo and BH mass. Coloured contours/lines refer to the analysis with different determinations of the local BH mass function: orange is for that obtained combining the local SMF with the $M_\bullet - M_\star$ relation, and magenta adds a narrow prior around $f_{\bullet \rightarrow \bullet} \approx 1$; blue is for that obtained combining the VDF with the observed $M_\bullet - \sigma$ relation; green is for that obtained combining the VDF with the debiased $M_\bullet - \sigma$ relation. The marginalized distributions are in arbitrary units (normalized to unity at their maximum value). Colored crosses mark bestfit values. In the uppermost panel, the gray shaded areas highlight non-physical values of the thin-disk radiative efficiency ϵ_{thin} which are excluded by hard priors.

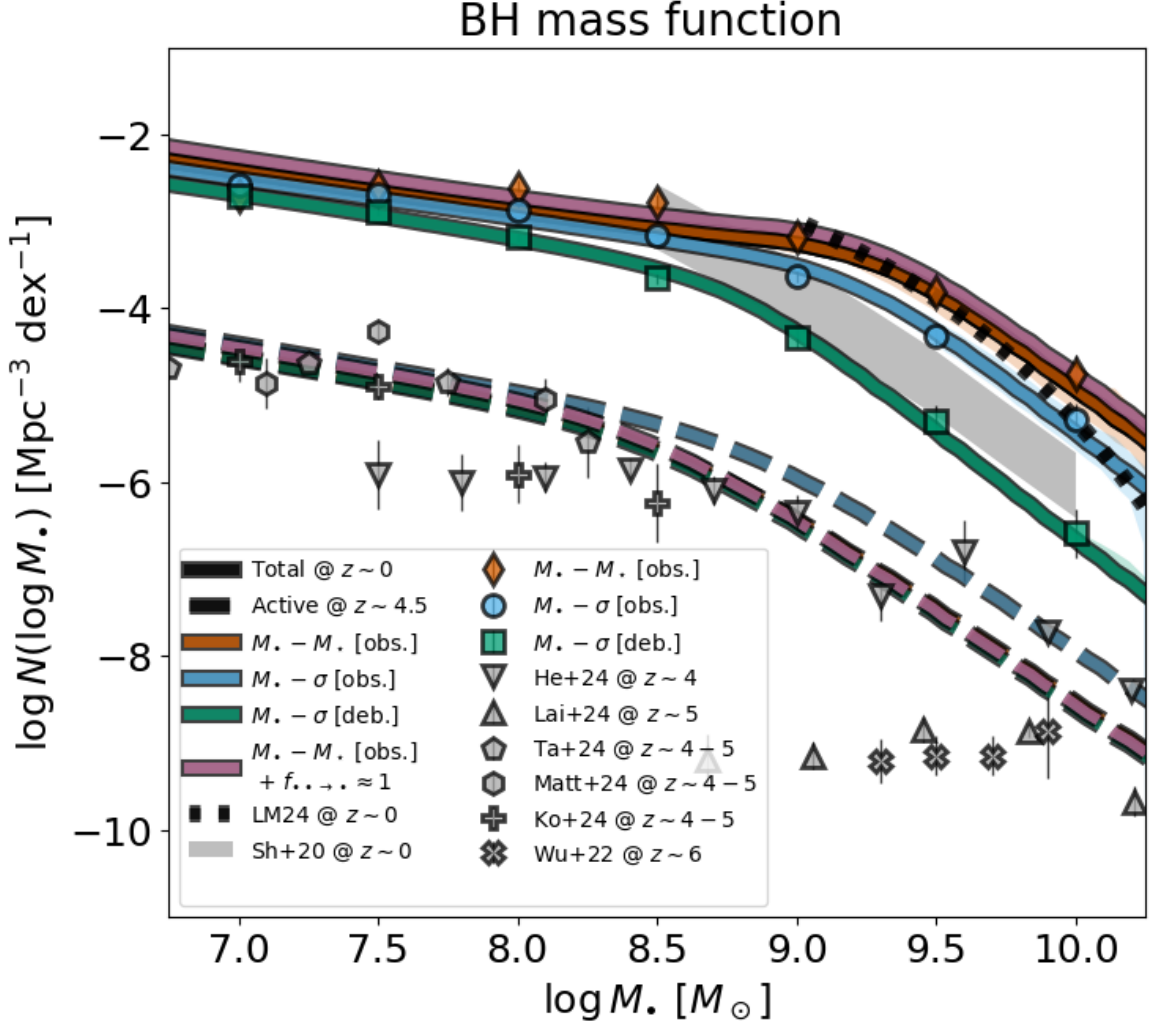


Figure 4. The supermassive BH mass function. Colored symbols refer to the local BH mass function determined from SMF and observed $M_\bullet - M_\star$ relation (green squares), from VDF and observed $M_\bullet - \sigma$ relation (blue circles), and from VDF and debiased $M_\bullet - \sigma$ relation (red rhomboids). These determinations bracket the overall systematics uncertainty on the estimates of the local BH mass function. For comparison, the dotted line is the recent determination based on the $M_\bullet - M_\star$ scaling relation by [183]; the gray shaded area is the determination by [79], obtained convolving the AGN luminosity function with the AGN luminosity vs. Eddington ratio relationship from SDSS [184, 185] and further correcting for obscured AGNs and inactive BHs. Grey symbols display recent estimates of the active BH mass function at $z \sim 4 - 6$ including BAL quasars and little red dots by [94] (reverse triangles), [95] (triangles), [203] (pentagons), [93] (hexagons), [96] (plus signs), and [73] (crosses). Solid lines with shaded areas illustrate the outcomes (median and 2σ credible intervals) for the local BH mass function reconstructed from our semi-empirical framework and MCMC analysis, with the same color-code as in previous figure, while dashed lines illustrate the corresponding active BH mass functions at $z \sim 4.5$.

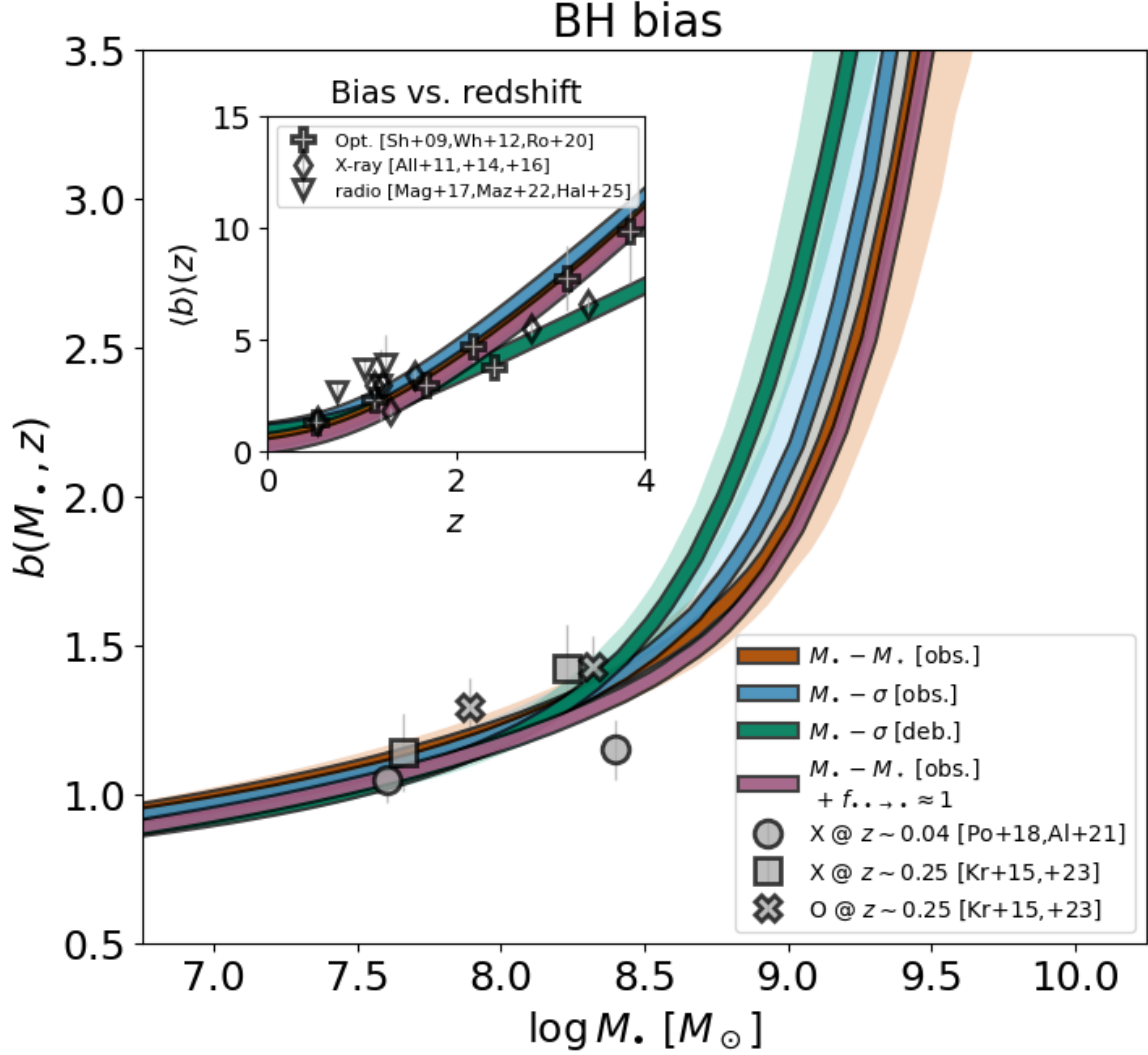


Figure 5. Large-scale bias as a function of BH mass. Data are from X-ray AGNs at $z \sim 0.04$ by [186, 187] (circles), from X-ray (squares) and optical (crosses) AGNs at $z \sim 0.25$ by [168, 169]. Colored solid lines with shaded areas illustrate the outcomes (median and 2σ credible intervals) from our semi-empirical framework model and MCMC analysis, with the same color code as in previous figures. The inset shows the BH bias as a function of redshift, averaged over the supermassive BH mass function. Colored lines are the outcomes from our semi-empirical framework, while symbols report observational estimates from optical (plus signs; [184, 189, 197]), X-ray (diamonds; [188, 191, 192]) and radio-selected (inverse triangles; [194, 199, 200]) samples. Note that the redshift dependence has not been fitted upon, and constitutes just an a-posteriori validation of our semi-empirical framework.

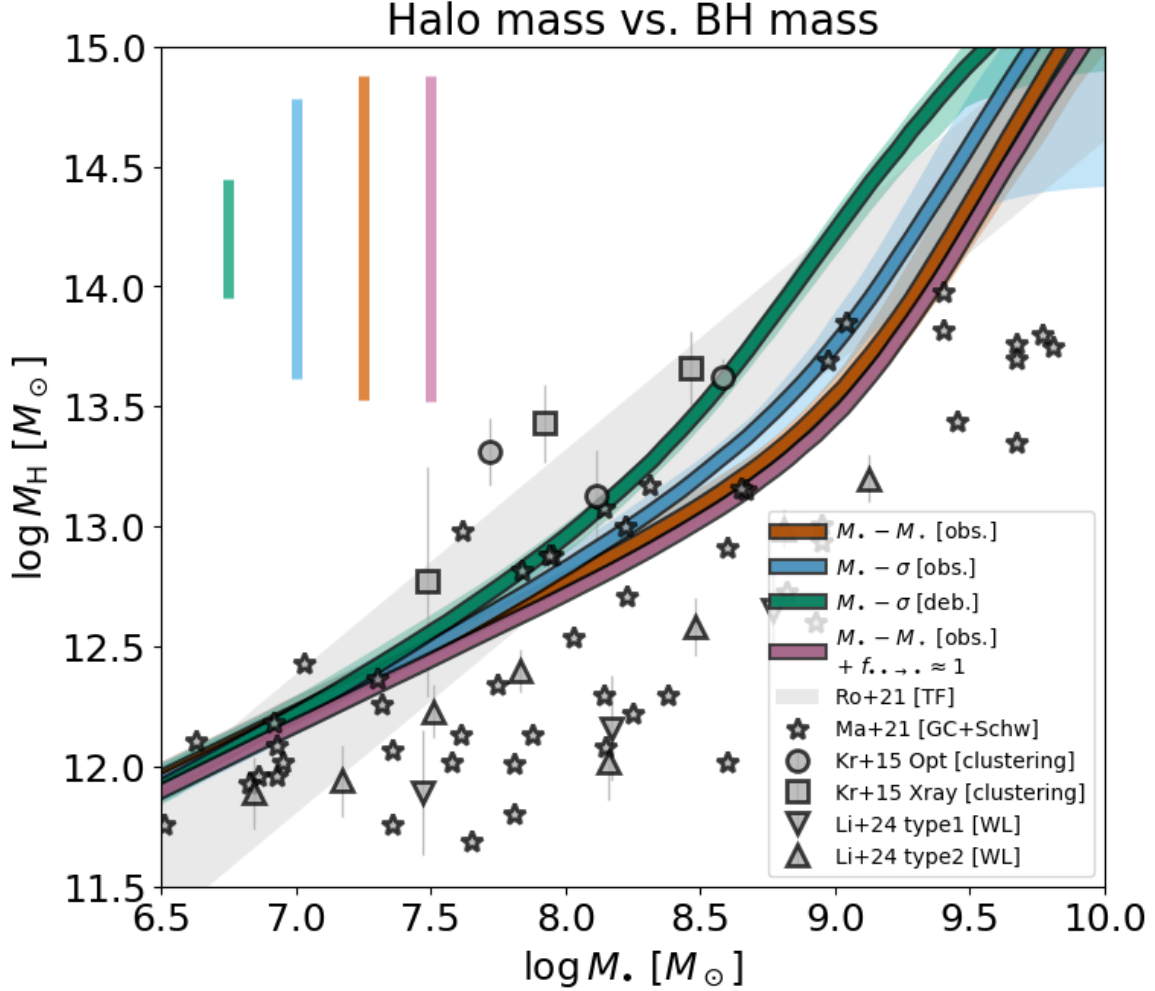


Figure 6. Halo mass vs. BH mass relationship. Data symbols refer to estimate from clustering of X-ray (squares) and optical (circles) AGNs by [168], from weak lensing measurements of type-1 (inverse triangles) and type-2 (triangles) AGNs, from globular cluster dynamics (stars) by [205], and from Tully-Fisher kinematics for reverberation-mapped AGNs by [206] (grey shaded area). Colored solid lines with shaded areas illustrate the outcomes (median and 2σ credible intervals) from our semi-empirical framework and MCMC analysis on BH demographics and clustering, with the same color code as in previous figures; note that the data reported here have not been fitted upon, and constitute just an a-posteriori validation. In the top left corner the colored errorbars represent the bestfit scatter $\sigma_{\log M_H}$ on the BH to halo mass relationship inferred from our analysis.

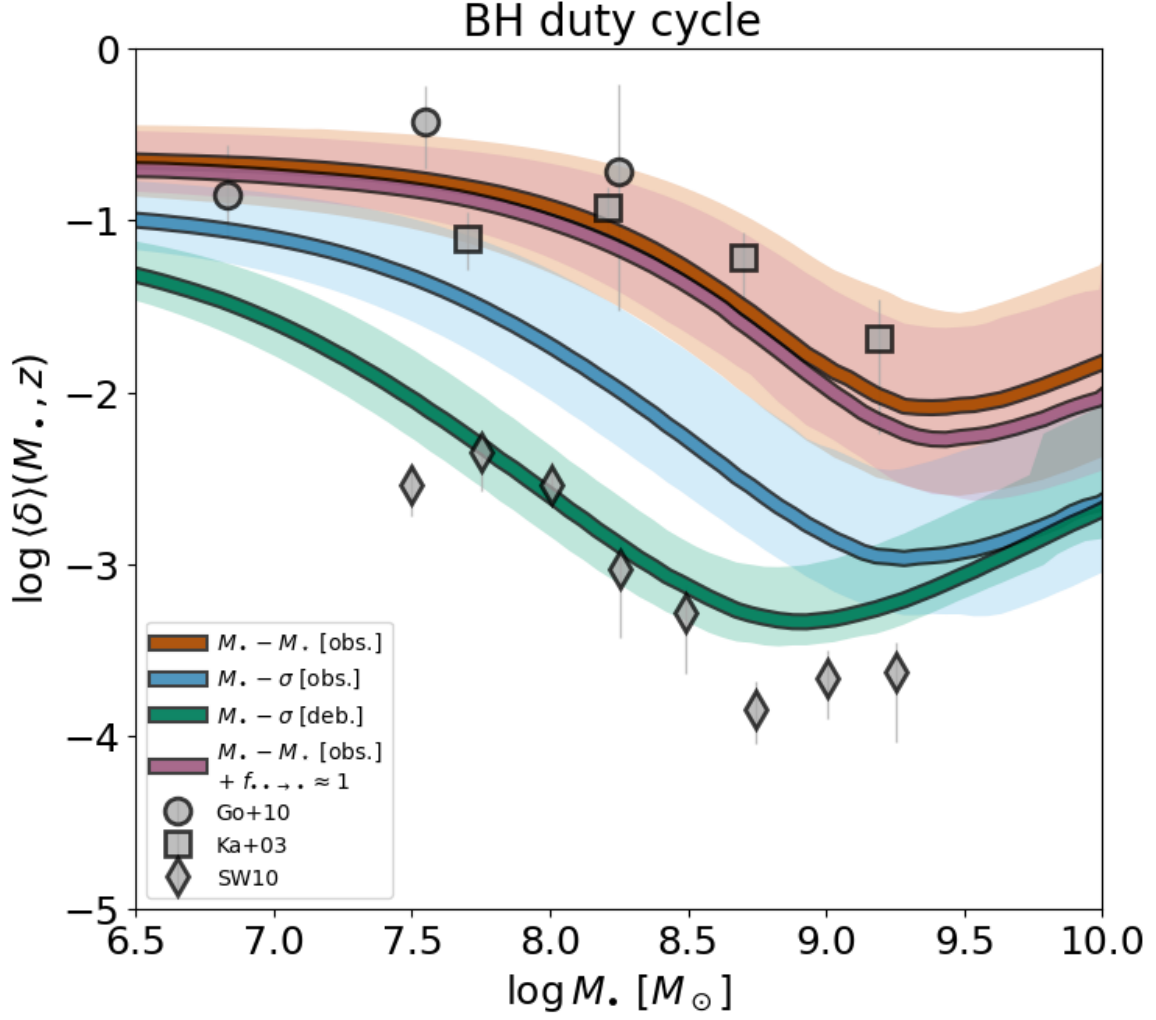


Figure 7. Local duty cycle as a function of BH mass. Data are from AGN active fractions by [202] (circles), [201] (squares) and [135] (diamonds). Colored solid lines with shaded areas illustrate the outcomes (median and 2σ credible intervals) from our semi-empirical framework and MCMC analysis, with the same color code as in previous figures; note that the data reported here have not been fitted upon, and constitute just an a-posteriori validation.

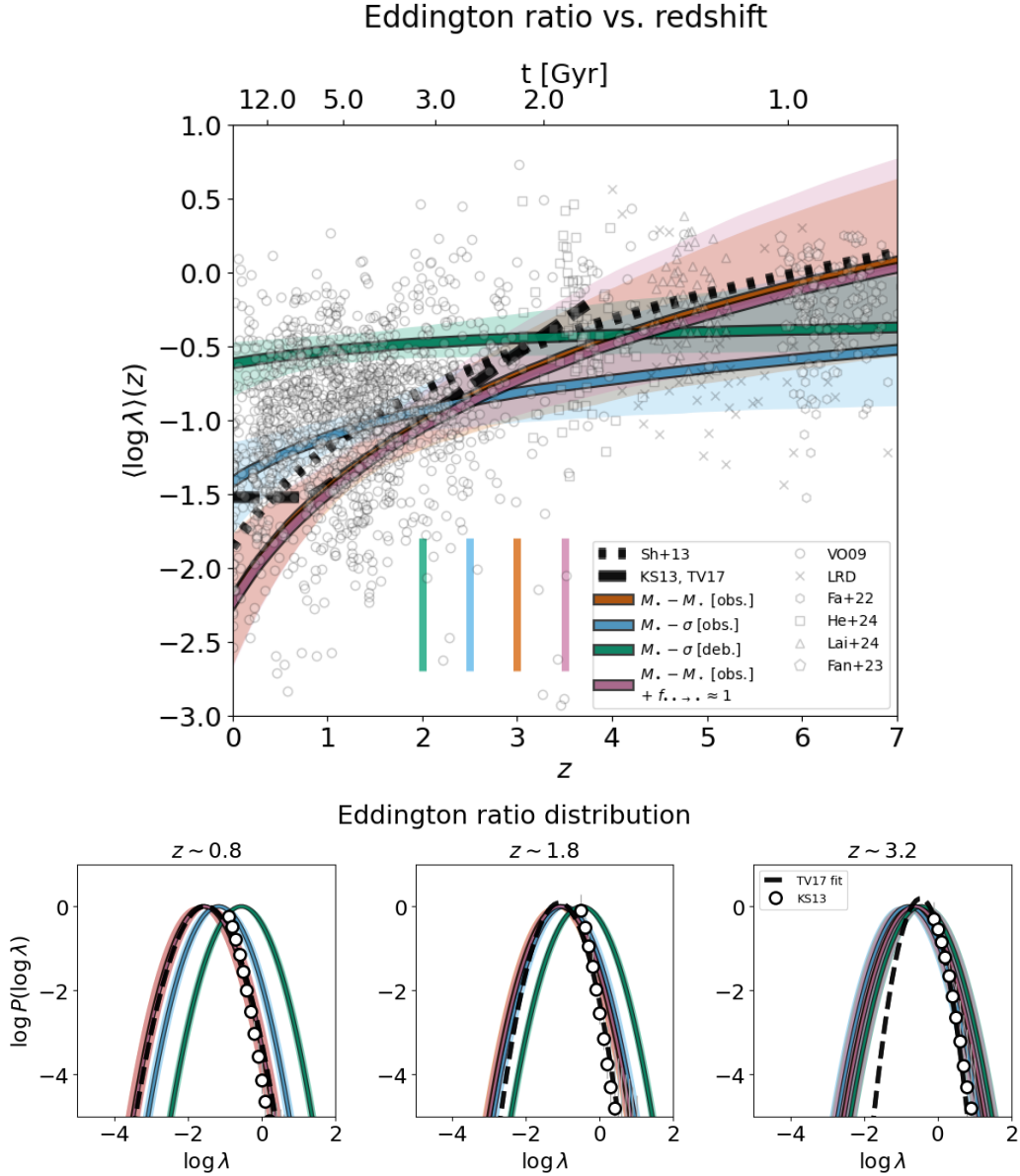


Figure 8. Top panel: average Eddington ratio vs. redshift (bottom axis) or cosmic time (top axis). Data are from: classic measurements from X-ray/mid-IR/optical AGNs by [70, 81, 185, 208] (circles), broad-absorption-line quasars by [94] (squares), [95] (triangles) [42] (diamonds) and [44] (pentagons); little red dots by [89, 93, 209, 210] (crosses). For reference, the dashed line illustrates the mean Eddington ratio found for type-1 AGNs at $z \lesssim 4$ by [72], as fitted by [31]; the dotted line is the dependence assumed on phenomenological grounds by [29]. Colored solid lines with shaded areas illustrate the outcomes (median and 2σ credible intervals) from our semi-empirical framework and MCMC analysis, with the same color code as in previous figures; note that the data reported here have not been fitted upon, and constitute just an a-posteriori validation. The colored errorbars in the bottom left corner represent the log-normal scatter in the Eddington ratio distribution adopted in our model. Bottom panels: Eddington ratio distribution at three representative redshifts $z \approx 0.8$ (left) 1.8 (middle) and 3.2 (right). Circles illustrate the estimates for type-1 AGNs by [72], plotted for the range in $\log \lambda$ where the completeness for the flux-limited SDSS sample is above 10%. Dashed line is the corresponding lognormal fit by [31]. Colored lines with shaded areas show the outcomes of our semi-empirical framework, as above.

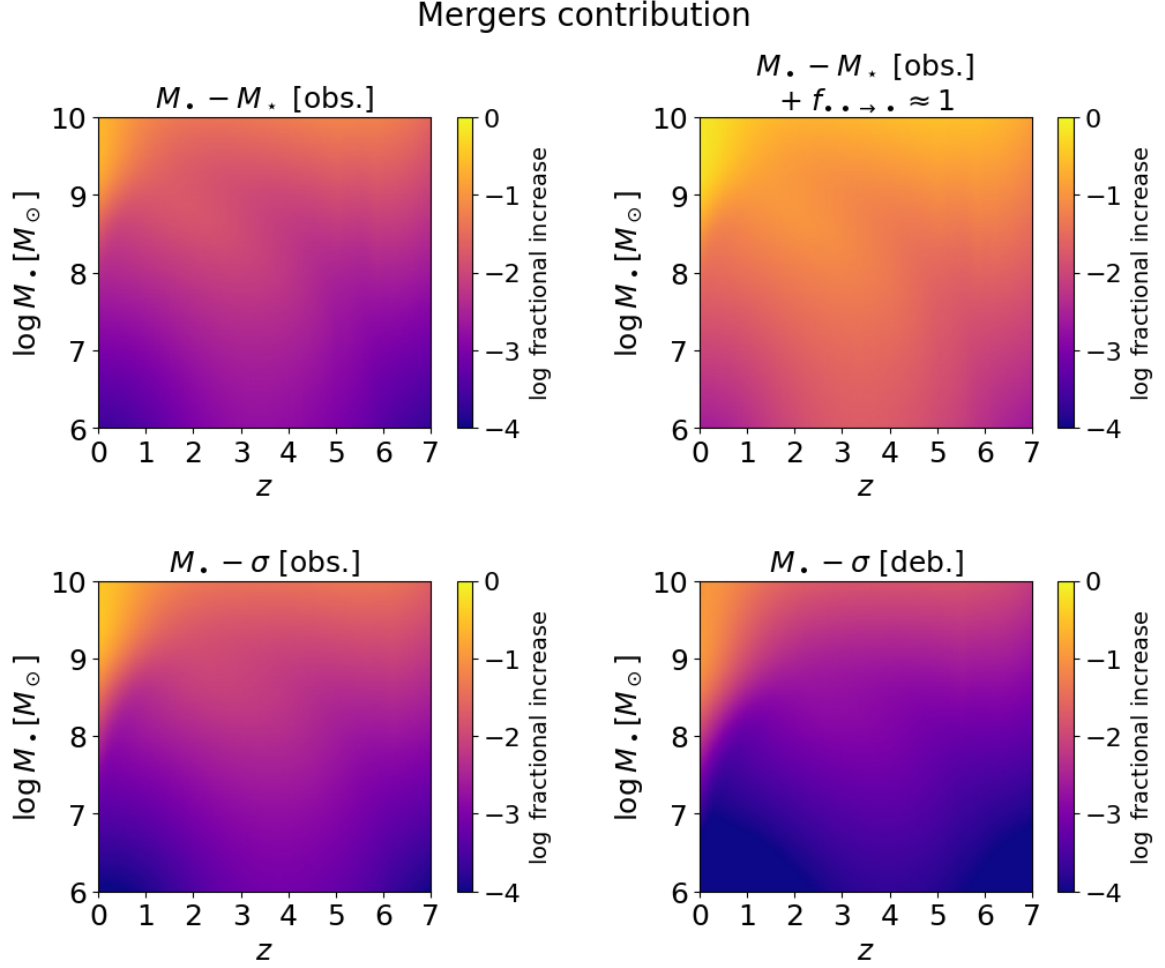


Figure 9. The relative contribution of binary BH mergers to the increase of the BH mass function in different mass and redshift bins (color-coded). Panels refer to the analysis with different determinations of the local BH mass function: top left is for SMF plus observed $M_{\bullet} - M_{\star}$; top right is for SMF plus observed $M_{\bullet} - M_{\star}$ plus a narrow prior centered on $f_{\bullet \rightarrow \bullet} \approx 1$; bottom left is for VDF plus observed $M_{\bullet} - \sigma$; bottom right is for VDF plus debiased $M_{\bullet} - \sigma$.

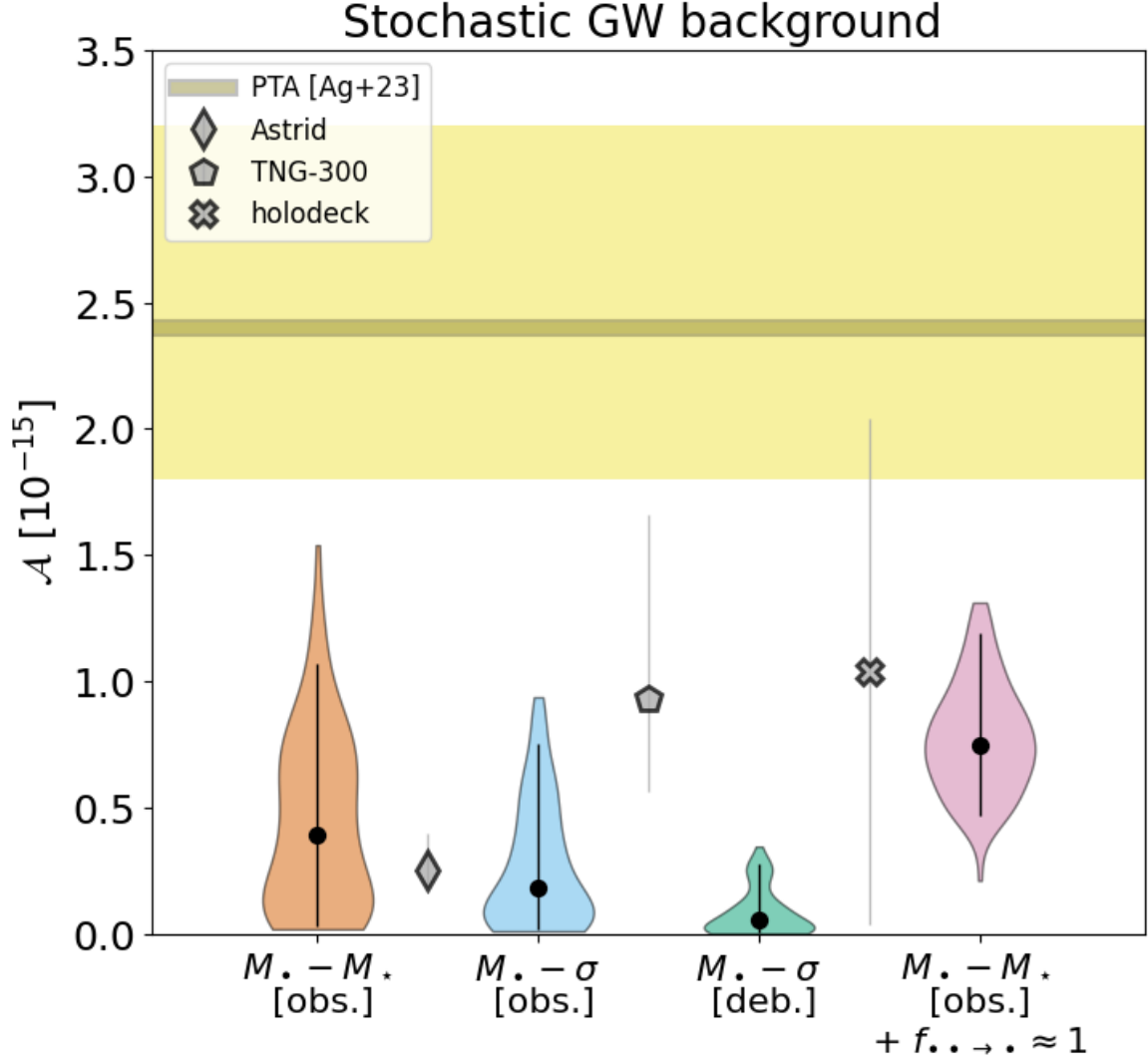


Figure 10. Normalized stretch amplitude of the gravitational wave background. Colored violins illustrate the full posterior distributions from our semi-empirical framework and MCMC analysis, with the same color code as in previous figures (and also specified on the bottom axis); the black dots and solid lines illustrate the median and the 2σ credible interval. Grey symbols display the results from the *Astrid* (rhomboid; [117]) and *TNG-300* numerical simulations (pentagon; [117]), and from the *holodeck* suite (cross; [101]). The measurement by the *NANOGrav* collaboration [101] is displayed by the yellow solid line with shaded area (median and 90% credible interval).

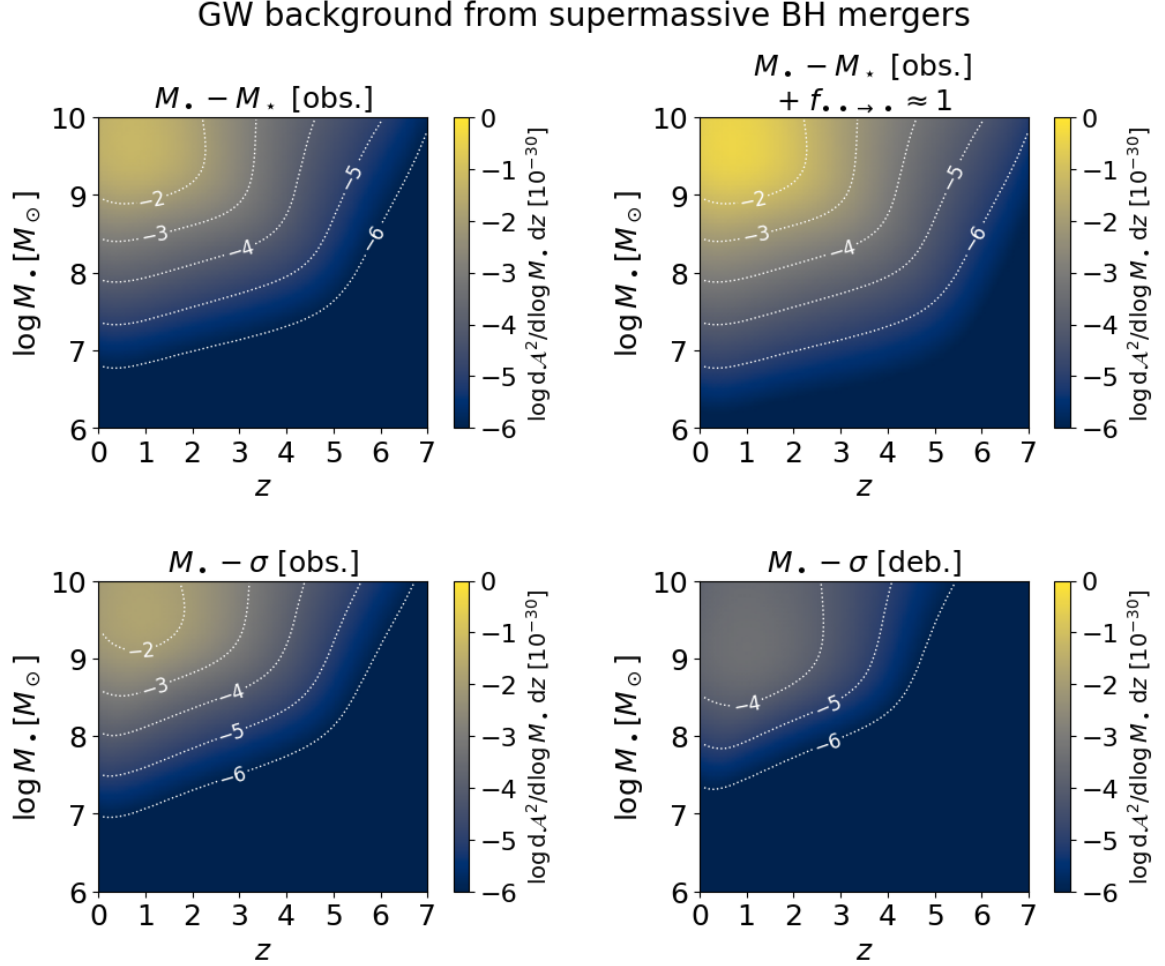


Figure 11. The contribution to the GW stochastic background from supermassive BH mergers sliced in redshift and BH mass. Panels refer to the analysis with different determination of the local BH mass function: top left is for SMF plus observed $M_\bullet - M_\star$; top right is for SMF plus observed $M_\bullet - M_\star$ plus a narrow prior centered on $f_{\bullet \rightarrow \bullet} \approx 1$; bottom left is for VDF plus observed $M_\bullet - \sigma$; bottom right is for VDF plus debiased $M_\bullet - \sigma$. The color-scale and white contours display logarithmic values of the normalized (squared) strain parameters $\mathcal{A}^2$ in bins of BH mass $\log M_\bullet$ and redshift z .

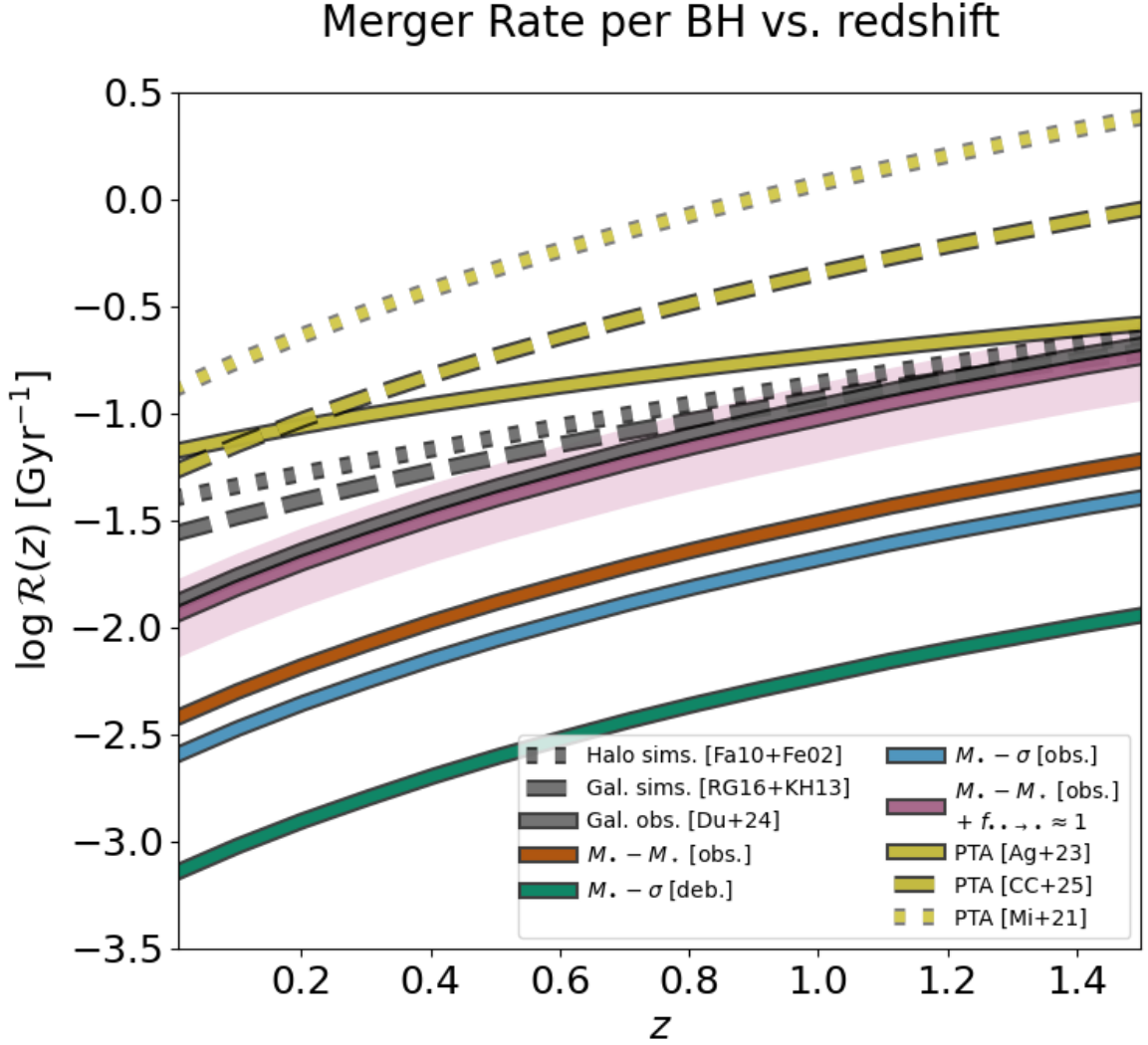


Figure 12. The merger rate per descendant BH vs. redshift, referred to a reference BH mass $M_{\bullet} \sim 10^{10} M_{\odot}$ as a representative contributor to the GW background. Colored solid lines illustrate the outcomes from our semi-empirical framework and MCMC analysis, with the same color code as in previous figures: the median from the posterior is reported, but the uncertainty for the red, blue and green cases essentially extends from the upper bound of the magenta line downwards (i.e., our model can only set upper limits on $f_{\bullet \rightarrow \bullet}$, see figure 3). The yellow lines are the reconstructed merger rate per BH from various analysis of PTA data: solid refers to [164], dashed to [220], and dotted to [218]. The grey lines are illustrated for comparison: solid is the major merger rate per galaxy estimated from JWST galaxy pairs counts by [162]; dashed is the major merger rate per galaxy from the *Illustris* simulation by [222], converted to the reference BH mass via the BH-stellar mass relation by [52]; dotted is the major merger rate per dark matter halo by [223], converted to the reference BH mass via the BH-halo mass relation by [173].

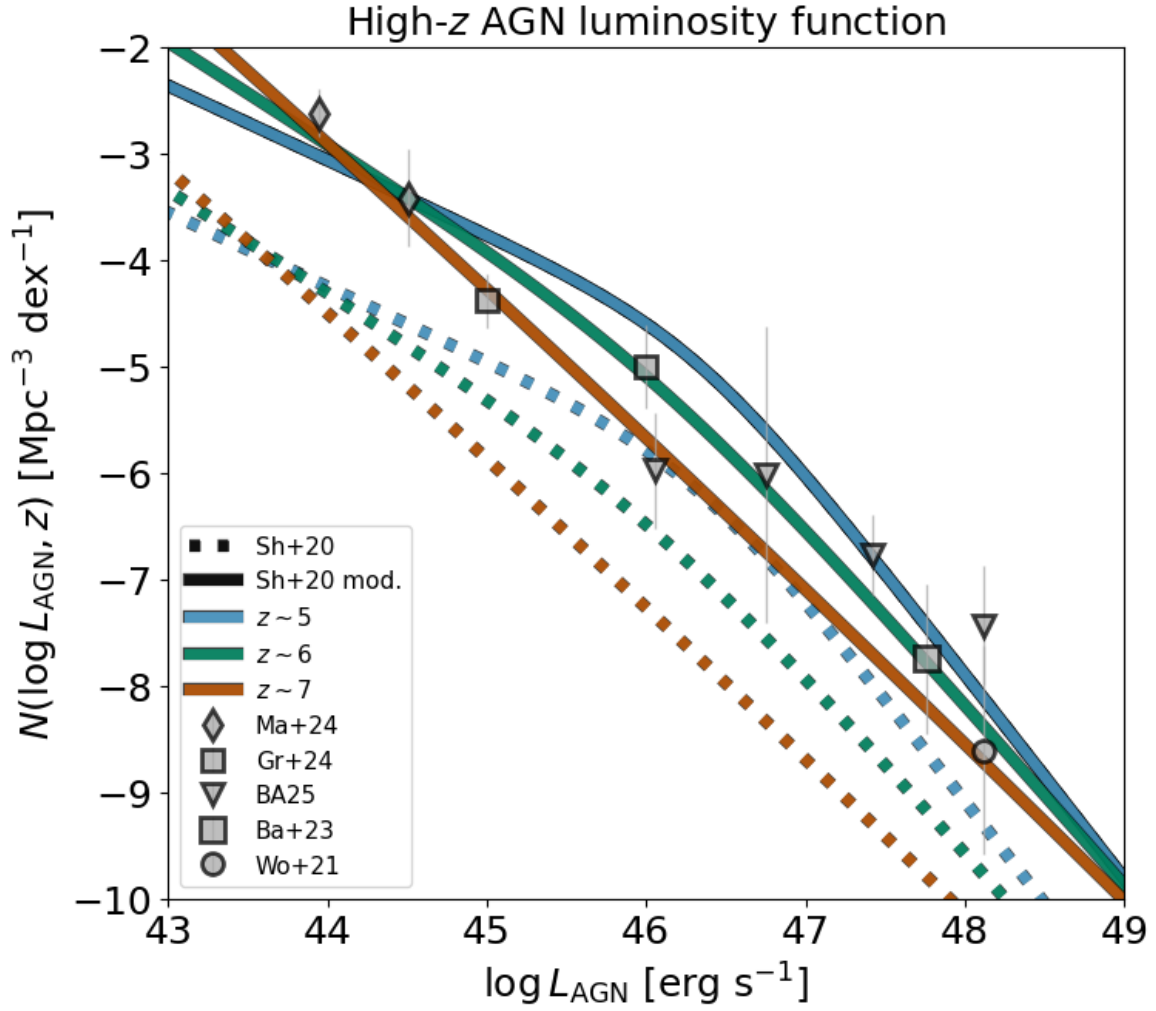


Figure 13. The AGN luminosity function at high redshift $z \gtrsim 4.5$. Colored lines refer to redshift $z \approx 5$ (blue), 6 (green), and 7 (red). Dotted lines represent the original luminosity function by [79] used as our fiducial input in the main text, while the solid lines are a modified version (see appendix A) to account for the recent data in the range $z \sim 5 - 7$ from [228] (diamonds), [209] (squares), [216] (squares), [217] (reversed triangles) and [227] (circle).

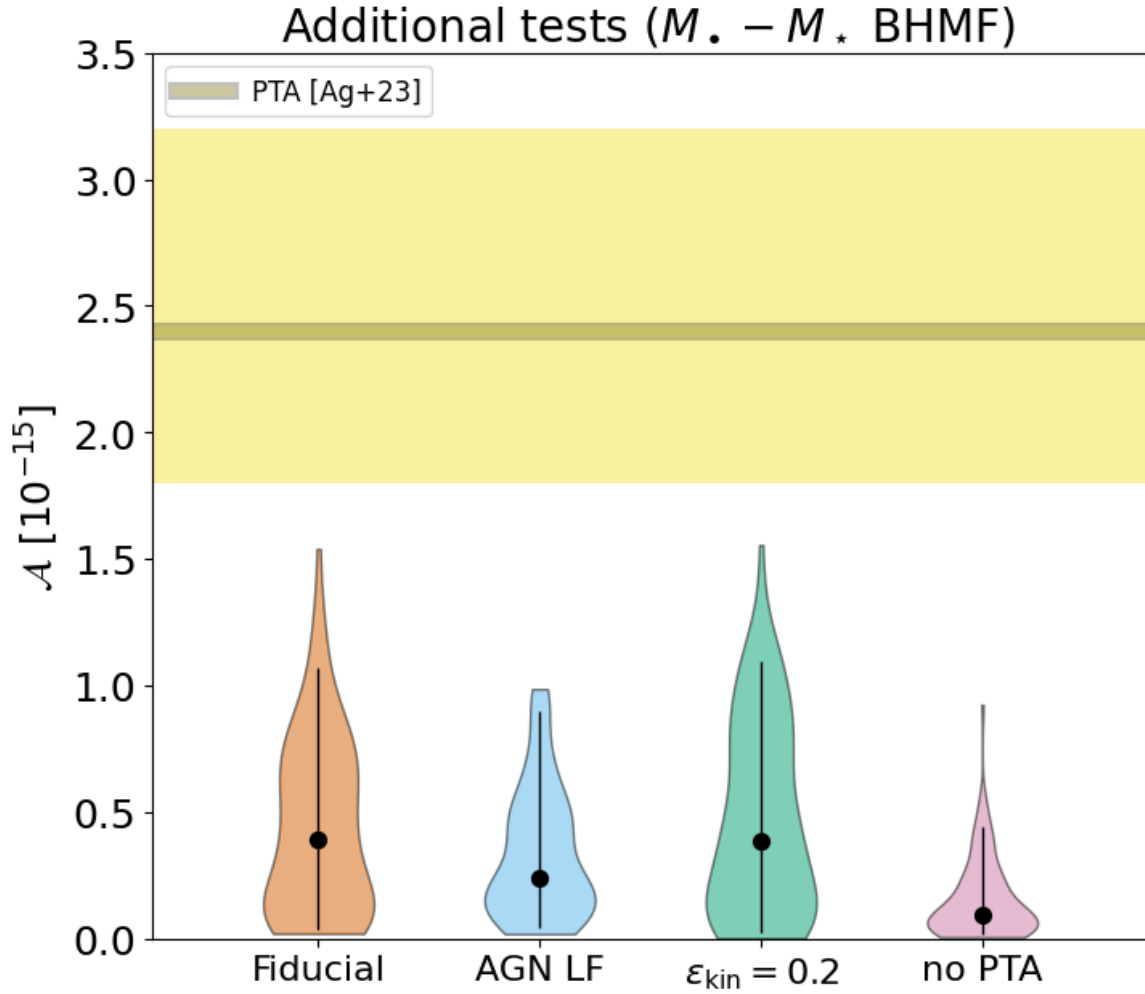


Figure 14. Additional tests on the contribution of binary BH mergers to the gravitational wave background. The local BH mass function determination from the stellar mass function and the $M_{\star} - M_{\bullet}$ relation is adopted. Colored violins illustrate the posterior distributions from our semi-empirical framework and MCMC analysis; the black dots and solid lines illustrate the median and the 2σ credible interval. The orange violin refers to our fiducial setting in the main text. The blue violin refers to the correction of the high- z AGN luminosity function discussed in appendix A. The green violin refers to a high value of the kinetic efficiency $\epsilon_{\text{kin}} = 0.2$. The magenta violin refers to the case when the PTA measurement of the GW background is not included in the likelihood. The measurement by the NANOGrav collaboration [101] is displayed by the yellow solid line with shaded area (median and 90% credible interval).

References

- [1] J. Magorrian et al., *The demography of massive dark objects in galaxy centers*, *Astron. J.* **115** (1998) 2285 [[astro-ph/9708072](#)] [[INSPIRE](#)].
- [2] L. Ferrarese and D. Merritt, *A fundamental relation between supermassive black holes and their host galaxies*, *Astrophys. J. Lett.* **539** (2000) L9 [[astro-ph/0006053](#)] [[INSPIRE](#)].

- [3] K. Gebhardt et al., *A relationship between nuclear black hole mass and galaxy velocity dispersion*, *Astrophys. J. Lett.* **539** (2000) L13 [[astro-ph/0006289](#)] [[INSPIRE](#)].
- [4] S. Tremaine et al., *The slope of the black hole mass versus velocity dispersion correlation*, *Astrophys. J.* **574** (2002) 740 [[astro-ph/0203468](#)] [[INSPIRE](#)].
- [5] C.J. Burke, P. Natarajan, V.F. Baldassare and M. Geha, *Multiwavelength constraints on the local black hole occupation fraction*, *Astrophys. J.* **978** (2025) 77 [[arXiv:2410.11177](#)] [[INSPIRE](#)].
- [6] F. Zou et al., *Central massive black holes are not ubiquitous in local low-mass galaxies*, *Astrophys. J.* **992** (2025) 176 [[arXiv:2510.05252](#)].
- [7] D. Lynden-Bell, *Galactic nuclei as collapsed old quasars*, *Nature* **223** (1969) 690 [[INSPIRE](#)].
- [8] N.I. Shakura and R.A. Sunyaev, *Black holes in binary systems. Observational appearance*, *Astron. Astrophys.* **24** (1973) 337 [[INSPIRE](#)].
- [9] EVENT HORIZON TELESCOPE collaboration, *First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole*, *Astrophys. J. Lett.* **875** (2019) L1 [[arXiv:1906.11238](#)] [[INSPIRE](#)].
- [10] EVENT HORIZON TELESCOPE collaboration, *First Sagittarius A* Event Horizon Telescope results. I. The shadow of the supermassive black hole in the center of the Milky Way*, *Astrophys. J. Lett.* **930** (2022) L12 [[arXiv:2311.08680](#)] [[INSPIRE](#)].
- [11] M. Hirschmann et al., *Cosmological simulations of black hole growth: AGN luminosities and downsizing*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 2304 [[arXiv:1308.0333](#)] [[INSPIRE](#)].
- [12] Y. Rosas-Guevara et al., *Supermassive black holes in the EAGLE universe. Revealing the observables of their growth*, *Mon. Not. Roy. Astron. Soc.* **462** (2016) 190 [[arXiv:1604.00020](#)].
- [13] M. Volonteri, Y. Dubois, C. Pichon and J. Devriendt, *The cosmic evolution of massive black holes in the horizon-AGN simulation*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 2979 [[arXiv:1602.01941](#)] [[INSPIRE](#)].
- [14] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [[arXiv:1707.03397](#)] [[INSPIRE](#)].
- [15] R. Weinberger et al., *Supermassive black holes and their feedback effects in the IllustrisTNG simulation*, *Mon. Not. Roy. Astron. Soc.* **479** (2018) 4056 [[arXiv:1710.04659](#)] [[INSPIRE](#)].
- [16] R. Davé et al., *Simba: cosmological simulations with black hole growth and feedback*, *Mon. Not. Roy. Astron. Soc.* **486** (2019) 2827 [[arXiv:1901.10203](#)] [[INSPIRE](#)].
- [17] Y. Ni et al., *The ASTRID simulation: the evolution of supermassive black holes*, *Mon. Not. Roy. Astron. Soc.* **513** (2022) 670 [[arXiv:2110.14154](#)] [[INSPIRE](#)].
- [18] S. Chakraborty et al., *Probing $z \gtrsim 6$ massive black holes with gravitational waves*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 758 [[arXiv:2305.09362](#)] [[INSPIRE](#)].
- [19] F. Zou et al., *The cosmic evolution of the supermassive black hole population: a hybrid observed accretion and simulated mergers approach*, *Astrophys. J.* **976** (2024) 6 [[arXiv:2409.16364](#)].
- [20] G. Kauffmann and M. Haehnelt, *A unified model for the evolution of galaxies and quasars*, *Mon. Not. Roy. Astron. Soc.* **311** (2000) 576 [[astro-ph/9906493](#)] [[INSPIRE](#)].
- [21] R.G. Bower et al., *The broken hierarchy of galaxy formation*, *Mon. Not. Roy. Astron. Soc.* **370** (2006) 645 [[astro-ph/0511338](#)] [[INSPIRE](#)].

- [22] D.J. Croton et al., *The many lives of AGN: cooling flows, black holes and the luminosities and colours of galaxies*, *Mon. Not. Roy. Astron. Soc.* **365** (2006) 11 [Erratum *ibid.* **367** (2006) 864] [[astro-ph/0602065](#)] [[INSPIRE](#)].
- [23] R.S. Somerville et al., *A semi-analytic model for the co-evolution of galaxies, black holes, and active galactic nuclei*, *Mon. Not. Roy. Astron. Soc.* **391** (2008) 481 [[arXiv:0808.1227](#)] [[INSPIRE](#)].
- [24] P. Dayal et al., *The hierarchical assembly of galaxies and black holes in the first billion years: predictions for the era of gravitational wave astronomy*, *Mon. Not. Roy. Astron. Soc.* **486** (2019) 2336 [[arXiv:1810.11033](#)] [[INSPIRE](#)].
- [25] F. Fontanot et al., *The rise of active galactic nuclei in the galaxy evolution and assembly semi-analytic model*, *Mon. Not. Roy. Astron. Soc.* **496** (2020) 3943 [[arXiv:2002.10576](#)] [[INSPIRE](#)].
- [26] R. Valiante et al., *Unveiling early black hole growth with multifrequency gravitational wave observations*, *Mon. Not. Roy. Astron. Soc.* **500** (2020) 4095 [[arXiv:2010.15096](#)] [[INSPIRE](#)].
- [27] S. Ghodla and C. Ilie, *Reconstructing PTA measurements via early seeding of supermassive black holes*, [arXiv:2507.06163](#) [[INSPIRE](#)].
- [28] P.F. Hopkins et al., *A unified, merger-driven model for the origin of starbursts, quasars, the cosmic X-ray background, supermassive black holes and galaxy spheroids*, *Astrophys. J. Suppl.* **163** (2006) 1 [[astro-ph/0506398](#)] [[INSPIRE](#)].
- [29] F. Shankar, D.H. Weinberg and J. Miralda-Escudé, *Accretion-driven evolution of black holes: Eddington ratios, duty cycles, and active galaxy fractions*, *Mon. Not. Roy. Astron. Soc.* **428** (2013) 421 [[arXiv:1111.3574](#)] [[INSPIRE](#)].
- [30] R. Aversa et al., *Black hole and galaxy coevolution from continuity equation and abundance matching*, *Astrophys. J.* **810** (2015) 74 [[arXiv:1507.07318](#)] [[INSPIRE](#)].
- [31] M. Tucci and M. Volonteri, *Constraining supermassive black hole evolution through the continuity equation*, *Astron. Astrophys.* **600** (2017) A64 [[arXiv:1603.00823](#)] [[INSPIRE](#)].
- [32] R. Carraro et al., *Coevolution of black hole accretion and star formation in galaxies up to $z = 3.5$* , *Astron. Astrophys.* **642** (2020) A65 [[arXiv:2007.11002](#)].
- [33] Zhang et al., *Trinity I: self-consistently modelling the dark matter halo–galaxy–supermassive black hole connection from $z = 0$ –10*, *Mon. Not. Roy. Astron. Soc.* **518** (2022) 2123 [Erratum *ibid.* **522** (2023) 3627] [[arXiv:2105.10474](#)] [[INSPIRE](#)].
- [34] M. Bosi et al., *StAGE: Stellar Archaeology-driven Galaxy Evolution*, *Astrophys. J.* **984** (2025) 117 [[arXiv:2503.22543](#)] [[INSPIRE](#)].
- [35] D.J. Mortlock et al., *A luminous quasar at a redshift of $z = 7.085$* , *Nature* **474** (2011) 616 [[arXiv:1106.6088](#)] [[INSPIRE](#)].
- [36] F. Wang et al., *An ultra-luminous quasar at $z = 5.363$ with a ten billion solar mass black hole and a Metal-Rich DLA at $z \sim 5$* , *Astrophys. J.* **807** (2015) L9 [[arXiv:1506.01765](#)].
- [37] F. Wang et al., *The discovery of a luminous broad absorption line quasar at a redshift of 7.02*, *Astrophys. J. Lett.* **869** (2018) L9 [[arXiv:1810.11925](#)].
- [38] E. Banados et al., *An 800-million-solar-mass black hole in a significantly neutral universe at redshift 7.5*, *Nature* **553** (2018) 473 [[arXiv:1712.01860](#)] [[INSPIRE](#)].
- [39] DES collaboration, *Three new VHS–DES quasars at $6.7 < z < 6.9$ and emission line properties at $z > 6.5$* , *Mon. Not. Roy. Astron. Soc.* **487** (2019) 1874 [[arXiv:1901.07456](#)] [[INSPIRE](#)].

- [40] Y. Matsuoka et al., *Subaru high- z exploration of low-luminosity quasars (SHELLQs). X. Discovery of 35 quasars and luminous galaxies at $5.7 \leq z \leq 7.0$* , *Astrophys. J.* **883** (2019) 183 [[arXiv:1908.07910](#)].
- [41] J. Yang et al., *Probing early supermassive black hole growth and quasar evolution with near-infrared spectroscopy of 37 reionization-era quasars at $6.3 < z \leq 7.64$* , *Astrophys. J.* **923** (2021) 262 [[arXiv:2109.13942](#)] [[INSPIRE](#)].
- [42] E.P. Farina et al., *The X-shooter/ALMA sample of quasars in the epoch of reionization. II. Black hole masses, Eddington ratios, and the formation of the first quasars*, *Astrophys. J.* **941** (2022) 106 [[arXiv:2207.05113](#)].
- [43] V. D’Odorico et al., *XQR-30: the ultimate XSHOOTER quasar sample at the reionization epoch*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 1399 [[arXiv:2305.05053](#)] [[INSPIRE](#)].
- [44] X. Fan, E. Banados and R.A. Simcoe, *Quasars and the intergalactic medium at cosmic dawn*, *Ann. Rev. Astron. Astrophys.* **61** (2023) 373 [[arXiv:2212.06907](#)] [[INSPIRE](#)].
- [45] P. Natarajan, *Seeds to monsters: tracing the growth of black holes in the universe*, *Gen. Rel. Grav.* **46** (2014) 1702 [[INSPIRE](#)].
- [46] P. Madau, F. Haardt and M. Dotti, *Super-critical growth of massive black holes from stellar-mass seeds*, *Astrophys. J. Lett.* **784** (2014) L38 [[arXiv:1402.6995](#)] [[INSPIRE](#)].
- [47] L. Mayer and S. Bonoli, *The route to massive black hole formation via merger-driven direct collapse: a review*, *Rept. Prog. Phys.* **82** (2019) 016901 [[arXiv:1803.06391](#)] [[INSPIRE](#)].
- [48] K. Inayoshi, E. Visbal and Z. Haiman, *The assembly of the first massive black holes*, *Ann. Rev. Astron. Astrophys.* **58** (2020) 27 [[arXiv:1911.05791](#)] [[INSPIRE](#)].
- [49] L. Boco, A. Lapi and L. Danese, *Growth of supermassive black hole seeds in ETG star-forming progenitors: multiple merging of stellar compact remnants via gaseous dynamical friction and gravitational wave emission*, *Astrophys. J.* **891** (2020) 94 [[arXiv:2002.03645](#)] [[INSPIRE](#)].
- [50] M. Volonteri, M. Habouzit and M. Colpi, *The origins of massive black holes*, *Nature Rev. Phys.* **3** (2021) 732 [[arXiv:2110.10175](#)] [[INSPIRE](#)].
- [51] F. Pacucci and A. Loeb, *The redshift evolution of the $M_{\bullet}$ - $M_{\star}$ relation for JWST’s supermassive black holes at $z > 4$* , *Astrophys. J.* **964** (2024) 154 [[arXiv:2401.04159](#)] [[INSPIRE](#)].
- [52] J. Kormendy and L.C. Ho, *Coevolution (or not) of supermassive black holes and host galaxies*, *Ann. Rev. Astron. Astrophys.* **51** (2013) 511 [[arXiv:1304.7762](#)] [[INSPIRE](#)].
- [53] N.J. McConnell and C.-P. Ma, *Revisiting the scaling relations of black hole masses and host galaxy properties*, *Astrophys. J.* **764** (2013) 184 [[arXiv:1211.2816](#)] [[INSPIRE](#)].
- [54] A.E. Reines and M. Volonteri, *Relations between central black hole mass and total galaxy stellar mass in the local universe*, *Astrophys. J.* **813** (2015) 82 [[arXiv:1508.06274](#)] [[INSPIRE](#)].
- [55] B.L. Davis, A.W. Graham and E. Cameron, *Black hole mass scaling relations for spiral galaxies. II. M_{BH} - $M_{\star, tot}$ and M_{BH} - $M_{\star, disk}$* , *Astrophys. J.* **869** (2018) 113 [[arXiv:1810.04888](#)].
- [56] N. Sahu, A.W. Graham and B.L. Davis, *Black hole mass scaling relations for early-type galaxies. I. M_{BH} - $M_{\star, sph}$ and M_{BH} - $M_{\star, gal}$* , *Astrophys. J.* **876** (2019) 155 [[arXiv:1903.04738](#)].
- [57] P. Zhu, L.C. Ho and H. Gao, *The correlation between black hole mass and stellar mass for classical bulges and the cores of ellipticals*, *Astrophys. J.* **907** (2021) 6 [[arXiv:2011.07216](#)].
- [58] M. Bernardi, R.K. Sheth, E. Tundo and J.B. Hyde, *Selection bias in the $M_{\bullet}$ - σ and $M_{\bullet}$ - L correlations and its consequences*, *Astrophys. J.* **660** (2007) 267 [[astro-ph/0609300](#)] [[INSPIRE](#)].

- [59] F. Shankar et al., *Selection bias in dynamically measured supermassive black hole samples: its consequences and the quest for the most fundamental relation*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 3119 [[arXiv:1603.01276](#)] [[INSPIRE](#)].
- [60] F. Shankar et al., *Black hole scaling relations of active and quiescent galaxies: addressing selection effects and constraining virial factors*, *Mon. Not. Roy. Astron. Soc.* **485** (2019) 1278 [[arXiv:1901.11036](#)] [[INSPIRE](#)].
- [61] F. Shankar et al., *Constraining black hole-galaxy scaling relations from the large-scale clustering of active galactic nuclei and implied mean radiative efficiency*, *Nature Astron.* **4** (2019) 282 [[arXiv:1910.10175](#)] [[INSPIRE](#)].
- [62] A. Soltan, *Masses of quasars*, *Mon. Not. Roy. Astron. Soc.* **200** (1982) 115 [[INSPIRE](#)].
- [63] P. Salucci, E. Szuszkiewicz, P. Monaco and L. Danese, *Mass function of dormant black holes and the evolution of the active galactic nuclei*, *Mon. Not. Roy. Astron. Soc.* **307** (1999) 637 [[astro-ph/9811102](#)] [[INSPIRE](#)].
- [64] S.W. Davis and A. Laor, *The radiative efficiency of accretion flows in individual AGN*, *Astrophys. J.* **728** (2011) 98 [[arXiv:1012.3213](#)] [[INSPIRE](#)].
- [65] S.I. Raimundo et al., *Can we measure the accretion efficiency of active galactic nuclei?*, *Mon. Not. Roy. Astron. Soc.* **419** (2012) 2529 [[arXiv:1109.6225](#)] [[INSPIRE](#)].
- [66] B. Trakhtenbrot, M. Volonteri and P. Natarajan, *On the accretion rates and radiative efficiencies of the highest redshift quasars*, *Astrophys. J. Lett.* **836** (2017) L1 [[arXiv:1611.00772](#)] [[INSPIRE](#)].
- [67] J. Aird, A.L. Coil and D.D. Kocevski, *AGN accretion and black hole growth across compact and extended galaxy evolution phases*, *Mon. Not. Roy. Astron. Soc.* **515** (2022) 4860 [[arXiv:2201.11756](#)] [[INSPIRE](#)].
- [68] B. Laloux et al., *Accretion properties of X-ray AGN: evidence for radiation-regulated obscuration with redshift-dependent host galaxy contribution*, *Mon. Not. Roy. Astron. Soc.* **532** (2024) 3459 [[arXiv:2403.07109](#)].
- [69] J.A. Kollmeier et al., *Black hole masses and Eddington ratios at $0.3 < z < 4$* , *Astrophys. J.* **648** (2006) 128 [[astro-ph/0508657](#)] [[INSPIRE](#)].
- [70] M. Vestergaard and P.S. Osmer, *Mass functions of the active black holes in distant quasars from the large bright quasar survey, the bright quasar survey, and the color-selected sample of the SDSS fall equatorial stripe*, *Astrophys. J.* **699** (2009) 800 [[arXiv:0904.3348](#)] [[INSPIRE](#)].
- [71] C.J. Willott et al., *Eddington-limited accretion and the black hole mass function at redshift 6*, *Astron. J.* **140** (2010) 546 [[arXiv:1006.1342](#)] [[INSPIRE](#)].
- [72] B.C. Kelly and Y. Shen, *The demographics of broad line quasars in the mass-luminosity plane II. Black hole mass and Eddington ratio functions*, *Astrophys. J.* **764** (2013) 45 [[arXiv:1209.0477](#)] [[INSPIRE](#)].
- [73] J. Wu et al., *Demographics of $z \sim 6$ quasars in the black hole mass–luminosity plane*, *Mon. Not. Roy. Astron. Soc.* **517** (2022) 2659 [[arXiv:2210.02518](#)] [[INSPIRE](#)].
- [74] A. Bongiorno et al., *AGN host galaxy mass function in COSMOS: is AGN feedback responsible for the mass-quenching of galaxies?*, *Astron. Astrophys.* **588** (2016) A78 [[arXiv:1601.02091](#)] [[INSPIRE](#)].

- [75] EBOSS collaboration, *Observational constraints on the specific accretion-rate distribution of X-ray-selected AGNs*, *Mon. Not. Roy. Astron. Soc.* **471** (2017) 1976 [[arXiv:1705.01133](#)] [[INSPIRE](#)].
- [76] J. Aird, A.L. Coil and A. Georgakakis, *X-rays across the galaxy population — II. The distribution of AGN accretion rates as a function of stellar mass and redshift*, *Mon. Not. Roy. Astron. Soc.* **474** (2018) 1225 [[arXiv:1705.01132](#)] [[INSPIRE](#)].
- [77] R. Carraro et al., *An Eddington ratio-driven origin for the L_X - M_* relation in quiescent and star-forming active galaxies*, *Mon. Not. Roy. Astron. Soc.* **512** (2022) 1185 [[arXiv:2202.07661](#)] [[INSPIRE](#)].
- [78] P.F. Hopkins, G.T. Richards and L. Hernquist, *An observational determination of the bolometric quasar luminosity function*, *Astrophys. J.* **654** (2007) 731 [[astro-ph/0605678](#)] [[INSPIRE](#)].
- [79] X. Shen et al., *The bolometric quasar luminosity function at $z = 0-7$* , *Mon. Not. Roy. Astron. Soc.* **495** (2020) 3252 [[arXiv:2001.02696](#)] [[INSPIRE](#)].
- [80] E. Lusso et al., *Bolometric luminosities and Eddington ratios of X-ray selected active galactic nuclei in the XMM-COSMOS survey*, *Mon. Not. Roy. Astron. Soc.* **425** (2012) 623 [[arXiv:1206.2642](#)] [[INSPIRE](#)].
- [81] F. Duras et al., *Universal bolometric corrections for active galactic nuclei over seven luminosity decades*, *Astron. Astrophys.* **636** (2020) A73 [[arXiv:2001.09984](#)].
- [82] Y. Ueda et al., *Toward the standard population synthesis model of the X-ray background: evolution of X-ray luminosity and absorption functions of active galactic nuclei including Compton-thick populations*, *Astrophys. J.* **786** (2014) 104 [[arXiv:1402.1836](#)] [[INSPIRE](#)].
- [83] J. Buchner et al., *Obscuration-dependent evolution of active galactic nuclei*, *Astrophys. J.* **802** (2015) 89 [[arXiv:1501.02805](#)] [[INSPIRE](#)].
- [84] T.T. Ananna et al., *The accretion history of AGNs. I. Supermassive black hole population synthesis model*, *Astrophys. J.* **871** (2019) 240 [[arXiv:1810.02298](#)] [[INSPIRE](#)].
- [85] B. Laloux et al., *The demographics of obscured AGN from X-ray spectroscopy guided by multiwavelength information*, *Mon. Not. Roy. Astron. Soc.* **518** (2022) 2546 [[arXiv:2211.03800](#)].
- [86] C. Andonie et al., *Obscuration beyond the nucleus: infrared quasars can be buried in extreme compact starbursts*, *Monthly Notices of the Royal Astronomical Society: Letters* **527** (2023) L144 [[arXiv:2310.02330](#)].
- [87] H. Übler et al., *GA-NIFS: a massive black hole in a low-metallicity AGN at $z \sim 5.55$ revealed by JWST/NIRSpec IFS*, *Astron. Astrophys.* **677** (2023) A145 [[arXiv:2302.06647](#)].
- [88] L.J. Furtak et al., *JWST UNCOVER: extremely red and compact object at $z_{\text{phot}} \sim 7.6$ triply imaged by A2744*, *Astrophys. J.* **952** (2023) 142 [[arXiv:2212.10531](#)].
- [89] Y. Harikane et al., *A JWST/NIRSpec first census of broad-line AGNs at $z = 4-7$: detection of 10 faint AGNs with $M_{\text{BH}} \sim 10^6-10^8 M_\odot$ and their host galaxy properties*, *Astrophys. J.* **959** (2023) 39 [[arXiv:2303.11946](#)].
- [90] V. Kokorev et al., *UNCOVER: a NIRSpec identification of a broad-line AGN at $z = 8.50$* , *Astrophys. J. Lett.* **957** (2023) L7 [[arXiv:2308.11610](#)].
- [91] M.A. Stone et al., *Undermassive host galaxies of five $z \sim 6$ luminous quasars detected with JWST*, *Astrophys. J.* **964** (2024) 90 [[arXiv:2310.18395](#)].

- [92] D.D. Kocevski et al., *Hidden little monsters: spectroscopic identification of low-mass, broad-line AGNs at $z > 5$ with CEERS*, *Astrophys. J. Lett.* **954** (2023) L4 [[arXiv:2302.00012](#)].
- [93] J. Matthee et al., *Little red dots: an abundant population of faint active galactic nuclei at $z \sim 5$ revealed by the EIGER and FRESCO JWST surveys*, *Astrophys. J.* **963** (2024) 129 [[arXiv:2306.05448](#)] [[INSPIRE](#)].
- [94] W. He et al., *Black hole mass and Eddington-ratio distributions of less-luminous quasars at $z \sim 4$ in the Subaru hyper supprime-cam wide field*, *Astrophys. J.* **962** (2024) 152 [[arXiv:2311.08922](#)].
- [95] S. Lai et al., *Supermassive black holes are growing slowly by $z \sim 5$* , *Mon. Not. Roy. Astron. Soc.* **531** (2024) 2245 [[arXiv:2405.10721](#)].
- [96] V. Kokorev et al., *A census of photometrically selected little red dots at $4 < z < 9$ in JWST blank fields*, *Astrophys. J.* **968** (2024) 38 [[arXiv:2401.09981](#)] [[INSPIRE](#)].
- [97] H.B. Akins et al., *COSMOS-Web: the overabundance and physical nature of “little red dots” — implications for early galaxy and SMBH assembly*, *Astrophys. J.* **991** (2025) 37 [[arXiv:2406.10341](#)] [[INSPIRE](#)].
- [98] J.M. Comerford, K. Schluns, J.E. Greene and R.J. Cool, *Dual supermassive black hole candidates in the AGN and galaxy evolution survey*, *Astrophys. J.* **777** (2013) 64 [[arXiv:1309.2284](#)] [[INSPIRE](#)].
- [99] J.D. Silverman et al., *Dual supermassive black holes at close separation revealed by the hyper Suprime-Cam Subaru strategic program*, *Astrophys. J.* **899** (2020) 154 [[arXiv:2007.05581](#)].
- [100] R.W. Pfeifle et al., *Super-size me. The big multi-AGN catalog (the big MAC) data release 1: the source catalog*, *Astrophys. J. Suppl.* **281** (2025) 25 [[arXiv:2411.12799](#)] [[INSPIRE](#)].
- [101] NANOGrav collaboration, *The NANOGrav 15 yr data set: evidence for a gravitational-wave background*, *Astrophys. J. Lett.* **951** (2023) L8 [[arXiv:2306.16213](#)] [[INSPIRE](#)].
- [102] EPTA and INPTA collaborations, *The second data release from the European Pulsar Timing Array — V. Search for continuous gravitational wave signals*, *Astron. Astrophys.* **690** (2024) A118 [[arXiv:2306.16226](#)] [[INSPIRE](#)].
- [103] D.J. Reardon et al., *Search for an isotropic gravitational-wave background with the Parkes pulsar timing array*, *Astrophys. J. Lett.* **951** (2023) L6 [[arXiv:2306.16215](#)] [[INSPIRE](#)].
- [104] H. Xu et al., *Searching for the nano-Hertz stochastic gravitational wave background with the Chinese pulsar timing array data release I*, *Res. Astron. Astrophys.* **23** (2023) 075024 [[arXiv:2306.16216](#)] [[INSPIRE](#)].
- [105] J.S.B. Wyithe and A. Loeb, *Low-frequency gravitational waves from massive black hole binaries: predictions for LISA and pulsar timing arrays*, *Astrophys. J.* **590** (2003) 691 [[astro-ph/0211556](#)] [[INSPIRE](#)].
- [106] A. Sesana, A. Vecchio and C.N. Colacino, *The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with pulsar timing arrays*, *Mon. Not. Roy. Astron. Soc.* **390** (2008) 192 [[arXiv:0804.4476](#)] [[INSPIRE](#)].
- [107] A. Kulier et al., *Understanding black hole mass assembly via accretion and mergers at late times in cosmological simulations*, *Astrophys. J.* **799** (2015) 178 [[arXiv:1307.3684](#)] [[INSPIRE](#)].
- [108] A. Sesana, F. Shankar, M. Bernardi and R.K. Sheth, *Selection bias in dynamically measured supermassive black hole samples: consequences for pulsar timing arrays*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) L6 [[arXiv:1603.09348](#)] [[INSPIRE](#)].

- [109] S. D’Ascoli et al., *Electromagnetic emission from supermassive binary black holes approaching merger*, *Astrophys. J.* **865** (2018) 140 [[arXiv:1806.05697](#)] [[INSPIRE](#)].
- [110] S. Chen, A. Sesana and C.J. Conselice, *Constraining astrophysical observables of galaxy and supermassive black hole binary mergers using pulsar timing arrays*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 401 [[arXiv:1810.04184](#)] [[INSPIRE](#)].
- [111] J. Simon, *Exploring proxies for the supermassive black hole mass function: implications for pulsar timing arrays*, *Astrophys. J. Lett.* **949** (2023) L24 [[arXiv:2306.01832](#)] [[INSPIRE](#)].
- [112] G. Sato-Polito, M. Zaldarriaga and E. Quataert, *Where are the supermassive black holes measured by PTAs?*, *Phys. Rev. D* **110** (2024) 063020 [[arXiv:2312.06756](#)] [[INSPIRE](#)].
- [113] G. Sato-Polito, M. Zaldarriaga and E. Quataert, *Evolution of supermassive black holes in light of PTA measurements: implications for growth by mergers and accretion*, *Phys. Rev. D* **112** (2025) 123018 [[arXiv:2501.09786](#)] [[INSPIRE](#)].
- [114] J. Salcido et al., *Music from the heavens — gravitational waves from supermassive black hole mergers in the EAGLE simulations*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 870 [[arXiv:1601.06156](#)] [[INSPIRE](#)].
- [115] L.Z. Kelley et al., *The gravitational wave background from massive black hole binaries in Illustris: spectral features and time to detection with pulsar timing arrays*, *Mon. Not. Roy. Astron. Soc.* **471** (2017) 4508 [[arXiv:1702.02180](#)] [[INSPIRE](#)].
- [116] M.S. Siwek, L.Z. Kelley and L. Hernquist, *The effect of differential accretion on the gravitational wave background and the present day MBH binary population*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 537 [[arXiv:2005.09010](#)] [[INSPIRE](#)].
- [117] N. Chen et al., *The gravitational-wave background from massive black holes in the ASTRID simulation*, *Astrophys. J. Lett.* **991** (2025) L19 [[arXiv:2502.01024](#)] [[INSPIRE](#)].
- [118] T. Naab and J.P. Ostriker, *Theoretical challenges in galaxy formation*, *Ann. Rev. Astron. Astrophys.* **55** (2017) 59 [[arXiv:1612.06891](#)].
- [119] R.S. Somerville and R. Davé, *Physical models of galaxy formation in a cosmological framework*, *Ann. Rev. Astron. Astrophys.* **53** (2015) 51 [[arXiv:1412.2712](#)] [[INSPIRE](#)].
- [120] A. Lapi, L. Boco and F. Shankar, *Semi-empirical models of galaxy formation and evolution*, in *Encyclopedia of astrophysics*, Elsevier, The Netherlands (2026), p. 554–575 [[DOI:10.1016/b978-0-443-21439-4.00060-2](#)].
- [121] L.T.C. Seeyave et al., *First Light and Reionisation Epoch Simulations (FLARES) XX: comparing semi-analytic models at high-redshift*, [arXiv:2511.11982](#).
- [122] M. Donnari et al., *Quenched fractions in the IllustrisTNG simulations: comparison with observations and other theoretical models*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 4760 [[arXiv:2008.00004](#)] [[INSPIRE](#)].
- [123] M. Habouzit et al., *Co-evolution of massive black holes and their host galaxies at high redshift: discrepancies from six cosmological simulations and the key role of JWST*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 3751 [[arXiv:2201.09892](#)] [[INSPIRE](#)].
- [124] R.A. Crain and F. van de Voort, *Hydrodynamical simulations of the galaxy population: enduring successes and outstanding challenges*, *Ann. Rev. Astron. Astrophys.* **61** (2023) 473 [[arXiv:2309.17075](#)] [[INSPIRE](#)].
- [125] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](#)] [[INSPIRE](#)].

- [126] A. Cavaliere, P. Morrison and K. Wood, *On quasar evolution*, *Astrophys. J.* **170** (1971) 223.
- [127] T.A. Small and R.D. Blandford, *Quasar evolution and the growth of black holes*, *Mon. Not. Roy. Astron. Soc.* **259** (1992) 725.
- [128] A. Marconi et al., *Local supermassive black holes, relics of active galactic nuclei and the x-ray background*, *Mon. Not. Roy. Astron. Soc.* **351** (2004) 169 [[astro-ph/0311619](#)] [[INSPIRE](#)].
- [129] F. Shankar et al., *Super-massive black hole demography: the match between the local and accreted mass functions*, *Mon. Not. Roy. Astron. Soc.* **354** (2004) 1020 [[astro-ph/0405585](#)] [[INSPIRE](#)].
- [130] Q. Yu and Y. Lu, *Toward precise constraints on growth of massive black holes*, *Astrophys. J.* **689** (2008) 732 [[arXiv:0808.3777](#)] [[INSPIRE](#)].
- [131] X. Cao, *Cosmological evolution of massive black holes: effects of Eddington ratio distribution and quasar lifetime*, *Astrophys. J.* **725** (2010) 388 [[arXiv:1010.0046](#)] [[INSPIRE](#)].
- [132] A. Sicilia et al., *The black hole mass function across cosmic time. II. Heavy seeds and (super)massive black holes*, *Astrophys. J.* **934** (2022) 66 [[arXiv:2206.07357](#)] [[INSPIRE](#)].
- [133] F. Shankar, D.H. Weinberg and J. Miralda-Escude, *Self-consistent models of the AGN and black hole populations: duty cycles, accretion rates, and the mean radiative efficiency*, *Astrophys. J.* **690** (2009) 20 [[arXiv:0710.4488](#)] [[INSPIRE](#)].
- [134] G. Kauffmann and T.M. Heckman, *Feast and famine: regulation of black hole growth in low redshift galaxies*, *Mon. Not. Roy. Astron. Soc.* **397** (2009) 135 [[arXiv:0812.1224](#)] [[INSPIRE](#)].
- [135] A. Schulze and L. Wisotzki, *Low redshift AGN in the Hamburg/ESO survey: II. The active black hole mass function and the distribution function of Eddington ratios*, *Astron. Astrophys.* **516** (2010) A87 [[arXiv:1004.2671](#)] [[INSPIRE](#)].
- [136] A.V. Alonso-Tetilla et al., *Characterizing the roles of transitory obscured phases and inner torus in shaping the fractions of obscured AGN at cosmic noon*, *Mon. Not. Roy. Astron. Soc.* (2025) [[arXiv:2509.04553](#)] [[INSPIRE](#)].
- [137] A. Lapi et al., *Quasar luminosity functions from joint evolution of black holes and host galaxies*, *Astrophys. J.* **650** (2006) 42 [[astro-ph/0603819](#)] [[INSPIRE](#)].
- [138] K.S. Thorne, *Disk accretion onto a black hole. 2. Evolution of the hole*, *Astrophys. J.* **191** (1974) 507 [[INSPIRE](#)].
- [139] M.A. Abramowicz, B. Czerny, J.P. Lasota and E. Szuszkiewicz, *Slim accretion disks*, *Astrophys. J.* **332** (1988) 646 [[INSPIRE](#)].
- [140] K.-Y. Watarai, J. Fukue, M. Takeuchi and S. Mineshige, *Galactic black-hole candidates shining at the Eddington luminosity*, *Publications of the Astronomical Society of Japan* **52** (2000) 133.
- [141] R. Narayan and I.-S. Yi, *Advection dominated accretion: selfsimilarity and bipolar outflows*, *Astrophys. J.* **444** (1995) 231 [[astro-ph/9411058](#)] [[INSPIRE](#)].
- [142] R.D. Blandford and M.C. Begelman, *Two dimensional adiabatic flows onto a black hole: 1. Fluid accretion*, *Mon. Not. Roy. Astron. Soc.* **349** (2004) 68 [[astro-ph/0306184](#)] [[INSPIRE](#)].
- [143] A. Merloni and S. Heinz, *A synthesis model for AGN evolution: supermassive black holes growth and feedback modes*, *Mon. Not. Roy. Astron. Soc.* **388** (2008) 1011 [[arXiv:0805.2499](#)] [[INSPIRE](#)].
- [144] F. Shankar, A. Cavaliere, M. Cirasuolo and L. Maraschi, *The optical-radio mapping: the kinetic efficiency of radio-loud AGNs*, *Astrophys. J.* **676** (2008) 131 [[arXiv:0712.3004](#)] [[INSPIRE](#)].

- [145] G. Ghisellini et al., *The role of relativistic jets in the heaviest and most active supermassive black holes at high redshift*, *Mon. Not. Roy. Astron. Soc.* **432** (2013) 2818 [[arXiv:1304.1152](#)] [[INSPIRE](#)].
- [146] K. Zubovas, *AGN must be very efficient at powering outflows*, *Mon. Not. Roy. Astron. Soc.* **479** (2018) 3189 [[arXiv:1806.08914](#)].
- [147] Q.-J. Yu and S. Tremaine, *Observational constraints on growth of massive black holes*, *Mon. Not. Roy. Astron. Soc.* **335** (2002) 965 [[astro-ph/0203082](#)] [[INSPIRE](#)].
- [148] V. Vittorini, F. Shankar and A. Cavaliere, *The impact of energy feedback on quasar evolution and black hole demographics*, *Mon. Not. Roy. Astron. Soc.* **363** (2005) 1376 [[astro-ph/0508640](#)] [[INSPIRE](#)].
- [149] F. Leyvraz, *Scaling theory and exactly solved models in the kinetics of irreversible aggregation*, *Phys. Rept.* **383** (2003) 95 [[cond-mat/0305670](#)].
- [150] M.V. Smoluchowski, *Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen* (in German), *Z. Phys.* **17** (1916) 557.
- [151] A. Cavaliere, S. Colafrancesco and N. Menci, *Merging in cosmic structures*, *Astrophys. J.* **392** (1992) 41.
- [152] A.J. Benson, M. Kamionkowski and S.H. Hassani, *Self-consistent theory of halo mergers*, *Mon. Not. Roy. Astron. Soc.* **357** (2005) 847 [[astro-ph/0407136](#)] [[INSPIRE](#)].
- [153] A.L. Erickcek, M. Kamionkowski and A.J. Benson, *Supermassive black hole merger rates: uncertainties from halo merger theory*, *Mon. Not. Roy. Astron. Soc.* **371** (2006) 1992 [[astro-ph/0604281](#)] [[INSPIRE](#)].
- [154] J. Ellis et al., *Gravitational waves from supermassive black hole binaries in light of the NANOGrav 15-year data*, *Phys. Rev. D* **109** (2024) L021302 [[arXiv:2306.17021](#)] [[INSPIRE](#)].
- [155] E. Neistein and A. Dekel, *Merger rates of dark-matter haloes*, *Mon. Not. Roy. Astron. Soc.* **388** (2008) 1792 [[arXiv:0802.0198](#)] [[INSPIRE](#)].
- [156] H. Koehn, A. Just, P. Berczik and M. Tremmel, *Dynamics of supermassive black hole triples in the ROMULUS25 cosmological simulation*, *Astron. Astrophys.* **678** (2023) A11 [[arXiv:2308.10894](#)] [[INSPIRE](#)].
- [157] M. Sayeb, L. Blecha and L.Z. Kelley, *MBH binary intruders: triple systems from cosmological simulations*, *Mon. Not. Roy. Astron. Soc.* **527** (2023) 7424 [[arXiv:2311.18228](#)] [[INSPIRE](#)].
- [158] A. Cavaliere, S. Colafrancesco and N. Menci, *The merging runaway*, *Astrophys. J.* **376** (1991) L37.
- [159] O. Fakhouri and C.-P. Ma, *Environmental dependence of dark matter halo growth I: halo merger rates*, *Mon. Not. Roy. Astron. Soc.* **394** (2009) 1825 [[arXiv:0808.2471](#)] [[INSPIRE](#)].
- [160] V. Rodriguez-Gomez et al., *The stellar mass assembly of galaxies in the Illustris simulation: growth by mergers and the spatial distribution of accreted stars*, *Mon. Not. Roy. Astron. Soc.* **458** (2016) 2371 [[arXiv:1511.08804](#)] [[INSPIRE](#)].
- [161] K.R.V. Casteels et al., *Galaxy And Mass Assembly (GAMA): refining the local galaxy merger rate using morphological information*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) 1157 [[arXiv:1409.1358](#)] [[INSPIRE](#)].
- [162] Q. Duan et al., *Galaxy mergers in the epoch of reionization — I. A JWST study of pair fractions, merger rates, and stellar mass accretion rates at $z = 4.5\text{--}11.5$* , *Mon. Not. Roy. Astron. Soc.* **540** (2025) 774 [[arXiv:2407.09472](#)].

- [163] D. Puskás et al., *Constraining the major merger history of $z \sim 3\text{--}9$ galaxies using JADES: dominant in situ star formation*, *Mon. Not. Roy. Astron. Soc.* **540** (2025) 2146.
- [164] NANOGrav collaboration, *The NANOGrav 15 yr data set: constraints on supermassive black hole binaries from the gravitational-wave background*, *Astrophys. J. Lett.* **952** (2023) L37 [[arXiv:2306.16220](#)] [[INSPIRE](#)].
- [165] M. Volonteri, *Formation of supermassive black holes*, *Astron. Astrophys. Rev.* **18** (2010) 279 [[arXiv:1003.4404](#)] [[INSPIRE](#)].
- [166] A. Sicilia et al., *The black hole mass function across cosmic times. I. Stellar black holes and light seed distribution*, *Astrophys. J.* **924** (2022) 56 [[arXiv:2110.15607](#)] [[INSPIRE](#)].
- [167] D. Roberts et al., *Too many or too massive? Investigating the high- z demography of active supermassive black holes from the James Webb Space Telescope*, *Mon. Not. Roy. Astron. Soc.* submitted (2025).
- [168] M. Krumpe et al., *The spatial clustering of ROSAT all-sky survey active galactic nuclei IV. More massive black holes reside in more massive dark matter halos*, *Astrophys. J.* **815** (2015) 21 [[arXiv:1509.01261](#)] [[INSPIRE](#)].
- [169] SDSS-IV TEAM and SPIDERS TEAM collaborations, *The spatial clustering of ROSAT all-sky survey active galactic nuclei. V. The evolution of broad-line AGN clustering properties in the last 6 Gyr*, *Astrophys. J.* **952** (2023) 109 [[arXiv:2304.02036](#)] [[INSPIRE](#)].
- [170] B.P. Moster, T. Naab and S.D.M. White, *Galactic star formation and accretion histories from matching galaxies to dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **428** (2013) 3121 [[arXiv:1205.5807](#)] [[INSPIRE](#)].
- [171] J.L. Tinker et al., *Toward a halo mass function for precision cosmology: the limits of universality*, *Astrophys. J.* **688** (2008) 709 [[arXiv:0803.2706](#)] [[INSPIRE](#)].
- [172] F. Shankar et al., *Probing the co-evolution of SMBHs and their hosts from scaling relations pairwise residuals: dominance of stellar velocity dispersion and host halo mass*, *Mon. Not. Roy. Astron. Soc.* **541** (2025) 2070 [[arXiv:2505.02920](#)] [[INSPIRE](#)].
- [173] L. Ferrarese, *Beyond the bulge: a fundamental relation between supermassive black holes and dark matter halos*, *Astrophys. J.* **578** (2002) 90 [[astro-ph/0203469](#)] [[INSPIRE](#)].
- [174] M.C. Powell et al., *BASS. XXXVI. Constraining the local supermassive black hole–halo connection with BASS DR2 AGNs*, *Astrophys. J.* **938** (2022) 77 [[arXiv:2209.02728](#)] [[INSPIRE](#)].
- [175] J.L. Tinker et al., *The large scale bias of dark matter halos: numerical calibration and model tests*, *Astrophys. J.* **724** (2010) 878 [[arXiv:1001.3162](#)] [[INSPIRE](#)].
- [176] B. Diemer, *COLOSSUS: a python toolkit for cosmology, large-scale structure, and dark matter halos*, *Astrophys. J. Suppl.* **239** (2018) 35 [[arXiv:1712.04512](#)] [[INSPIRE](#)].
- [177] E.S. Phinney, *A practical theorem on gravitational wave backgrounds*, [astro-ph/0108028](#) [[INSPIRE](#)].
- [178] F.A. Jenet et al., *Upper bounds on the low-frequency stochastic gravitational wave background from pulsar timing observations: current limits and future prospects*, *Astrophys. J.* **653** (2006) 1571 [[astro-ph/0609013](#)] [[INSPIRE](#)].
- [179] F. Fastidio et al., *Realistic consecutive galaxy mergers form eccentric pulsar timing array sources*, *Astron. Astrophys.* **703** (2025) A86 [[arXiv:2506.16539](#)] [[INSPIRE](#)].
- [180] A.J. Moffett et al., *Galaxy And Mass Assembly (GAMA): the stellar mass budget by galaxy type*, *Mon. Not. Roy. Astron. Soc.* **457** (2016) 1308 [[arXiv:1512.02342](#)].

- [181] G.A.D. Savorgnan, A.W. Graham, A. Marconi and E. Sani, *Supermassive black holes and their host spheroids. II. The red and blue sequence in the M_{BH} - $M_{*,sph}$ diagram*, *Astrophys. J.* **817** (2016) 21 [[arXiv:1511.07437](#)].
- [182] SDSS collaboration, *The velocity dispersion function of early-type galaxies*, *Astrophys. J.* **594** (2003) 225 [[astro-ph/0303092](#)] [[INSPIRE](#)].
- [183] E.R. Liepold and C.-P. Ma, *Big galaxies and big black holes: the massive ends of the local stellar and black hole mass functions and the implications for nanoHertz gravitational waves*, *Astrophys. J. Lett.* **971** (2024) L29 [[arXiv:2407.14595](#)] [[INSPIRE](#)].
- [184] Y. Shen et al., *Quasar clustering from SDSS DR5: dependences on physical properties*, *Astrophys. J.* **697** (2009) 1656 [[arXiv:0810.4144](#)] [[INSPIRE](#)].
- [185] K. Nobuta et al., *Black hole mass and Eddington ratio distribution functions of X-ray selected broad-line AGNs at $z \sim 1.4$ in the Subaru XMM-Newton deep field*, *Astrophys. J.* **761** (2012) 143 [[arXiv:1211.0069](#)] [[INSPIRE](#)].
- [186] M.C. Powell et al., *The Swift/BAT AGN spectroscopic survey. IX. The clustering environments of an unbiased sample of local AGNs*, *Astrophys. J.* **858** (2018) 110 [[arXiv:1803.07589](#)].
- [187] V. Allevato et al., *Building robust active galactic nuclei mock catalogs to unveil black hole evolution and for survey planning*, *Astrophys. J.* **916** (2021) 34 [[arXiv:2105.02883](#)] [[INSPIRE](#)].
- [188] V. Allevato et al., *The XMM-Newton wide field survey in the COSMOS field: redshift evolution of AGN bias and subdominant role of mergers in triggering moderate luminosity AGN at redshift up to 2.2*, *Astrophys. J.* **736** (2011) 99 [[arXiv:1105.0520](#)] [[INSPIRE](#)].
- [189] BOSS collaboration, *The clustering of intermediate redshift quasars as measured by the baryon oscillation spectroscopic survey*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 933 [[arXiv:1203.5306](#)] [[INSPIRE](#)].
- [190] BOSS collaboration, *Cross-correlation of SDSS DR7 quasars and DR10 BOSS galaxies: the weak luminosity dependence of quasar clustering at $z \sim 0.5$* , *Astrophys. J.* **778** (2013) 98 [[arXiv:1212.4526](#)] [[INSPIRE](#)].
- [191] V. Allevato et al., *Clustering of moderate luminosity X-ray-selected type 1 and type 2 AGNs at $z \sim 3$* , *Astrophys. J.* **796** (2014) 4 [[arXiv:1409.7693](#)] [[INSPIRE](#)].
- [192] V. Allevato et al., *The Chandra cosmos legacy survey: clustering of X-ray-selected AGNs at $2.9 \leq z \leq 5.5$ using photometric redshift probability distribution functions*, *Astrophys. J.* **832** (2016) 70 [[arXiv:1609.02553](#)].
- [193] G. Mountrichas et al., *The clustering amplitude of X-ray-selected AGN at $z \sim 0.8$: evidence for a negative dependence on accretion luminosity*, *Mon. Not. Roy. Astron. Soc.* **457** (2016) 4195 [[arXiv:1602.01856](#)] [[INSPIRE](#)].
- [194] M. Magliocchetti et al., *The clustering properties of radio-selected AGN and star-forming galaxies up to redshifts $z \sim 3$* , *Mon. Not. Roy. Astron. Soc.* **464** (2017) 3271 [[arXiv:1606.08286](#)] [[INSPIRE](#)].
- [195] C.L. Hale et al., *The clustering and bias of radio-selected AGN and star-forming galaxies in the COSMOS field*, *Mon. Not. Roy. Astron. Soc.* **474** (2018) 4133 [[arXiv:1711.05201](#)] [[INSPIRE](#)].
- [196] S. Eftekharzadeh, A.D. Myers and E. Kourkchi, *A halo occupation interpretation of quasars at $z \sim 1.5$ using very small-scale clustering information*, *Mon. Not. Roy. Astron. Soc.* **486** (2019) 274 [[arXiv:1812.05760](#)] [[INSPIRE](#)].

- [197] EBOSS collaboration, *The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: large-scale structure catalogues for cosmological analysis*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 2354 [[arXiv:2007.09000](#)] [[INSPIRE](#)].
- [198] C. Krishnan et al., *The clustering of X-ray AGN at $0.5 < z < 4.5$: host galaxies dictate dark matter halo mass*, *Mon. Not. Roy. Astron. Soc.* **494** (2020) 1693 [[arXiv:2003.10461](#)].
- [199] A. Mazumder, A. Chakraborty and A. Datta, *A study on the clustering properties of radio-selected sources in the Lockman hole region at 325 MHz*, *Mon. Not. Roy. Astron. Soc.* **517** (2022) 3407 [[arXiv:2208.00992](#)] [[INSPIRE](#)].
- [200] C.L. Hale et al., *The clustering of active galactic nuclei and star forming galaxies in the LoTSS DeepFields*, *Mon. Not. Roy. Astron. Soc.* **544** (2025) 1323 [[arXiv:2510.01029](#)] [[INSPIRE](#)].
- [201] SDSS collaboration, *The host galaxies of AGN*, *Mon. Not. Roy. Astron. Soc.* **346** (2003) 1055 [[astro-ph/0304239](#)] [[INSPIRE](#)].
- [202] A.D. Goulding, D.M. Alexander, B.D. Lehmer and J.R. Mullaney, *Towards a complete census of AGNs in nearby galaxies: the incidence of growing black holes*, *Mon. Not. Roy. Astron. Soc.* **406** (2010) 597 [[arXiv:1003.3015](#)] [[INSPIRE](#)].
- [203] A.J. Taylor et al., *Broad-line AGNs at $3.5 < z < 6$: the black hole mass function and a connection with little red dots*, *Astrophys. J.* **986** (2025) 165 [[arXiv:2409.06772](#)].
- [204] S. Geris et al., *JADES reveals a large population of low-mass black holes at high redshift*, *Mon. Not. Roy. Astron. Soc.* **545** (2025) [[arXiv:2506.22147](#)].
- [205] A. Marasco et al., *A universal relation between the properties of supermassive black holes, galaxies, and dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 4274 [[arXiv:2105.10508](#)] [[INSPIRE](#)].
- [206] J.H. Robinson et al., *Tully-Fisher distances and dynamical mass constraints for 24 host galaxies of reverberation-mapped AGNs*, *Astrophys. J.* **912** (2021) 160 [[arXiv:2103.07000](#)].
- [207] Q. Li et al., *Black hole–halo mass relation from UNIONS weak lensing*, *Astrophys. J. Lett.* **969** (2024) L25 [[arXiv:2402.10740](#)] [[INSPIRE](#)].
- [208] Y.S. Dai et al., *Mid-infrared selected quasars I: virial black hole mass and Eddington ratios*, *Astrophys. J.* **791** (2014) 113 [[arXiv:1406.4477](#)] [[INSPIRE](#)].
- [209] J.E. Greene et al., *UNCOVER spectroscopy confirms the surprising ubiquity of active galactic nuclei in red sources at $z > 5$* , *Astrophys. J.* **964** (2024) 39 [[arXiv:2309.05714](#)] [[INSPIRE](#)].
- [210] R. Maiolino et al., *A small and vigorous black hole in the early universe*, *Nature* **627** (2024) 59 [*Erratum ibid.* **630** (2024) E2] [[arXiv:2305.12492](#)] [[INSPIRE](#)].
- [211] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: the MCMC hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306 [[arXiv:1202.3665](#)] [[INSPIRE](#)].
- [212] J.M. Bardeen, W.H. Press and S.A. Teukolsky, *Rotating black holes: locally nonrotating frames, energy extraction, and scalar synchrotron radiation*, *Astrophys. J.* **178** (1972) 347 [[INSPIRE](#)].
- [213] X. Zhang and Y. Lu, *On constraining the growth history of massive black holes via their distribution on the spin–mass plane*, *Astrophys. J.* **873** (2019) 101 [[arXiv:1902.07056](#)] [[INSPIRE](#)].
- [214] E. Giallongo et al., *Space densities and emissivities of active galactic nuclei at $z > 4$* , *Astrophys. J.* **884** (2019) 19.
- [215] G. Kulkarni, G. Worseck and J.F. Hennawi, *Evolution of the AGN UV luminosity function from redshift 7.5*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 1035 [[arXiv:1807.09774](#)] [[INSPIRE](#)].

- [216] C.L. Barlow-Hall et al., *Constraints on the X-ray luminosity function of AGN at $z = 5.7\text{--}6.4$ with the extragalactic serendipitous Swift survey*, *Mon. Not. Roy. Astron. Soc.* **519** (2023) 6055 [[arXiv:2201.11139](#)] [[INSPIRE](#)].
- [217] C.L. Barlow-Hall and J. Aird, *Measurements of the $z = 4\text{--}10$ X-ray luminosity function: the high space density of moderate-luminosity, obscured AGN*, [arXiv:2506.16145](#) [[INSPIRE](#)].
- [218] H. Middleton et al., *Retracted: correction to: massive black hole binary systems and the NANOGrav 12.5 yr results*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) L99 [Erratum *ibid.* **526** (2023) L34] [[arXiv:2011.01246](#)] [[INSPIRE](#)].
- [219] D. Izquierdo-Villalba, A. Sesana, S. Bonoli and M. Colpi, *Massive black hole evolution models confronting the n-Hz amplitude of the stochastic gravitational wave background*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 3488 [[arXiv:2108.11671](#)] [[INSPIRE](#)].
- [220] J.A. Casey-Clyde et al., *Quasars can signpost supermassive black hole binaries*, *Astrophys. J.* **987** (2025) 106 [[arXiv:2405.19406](#)] [[INSPIRE](#)].
- [221] Á. Kis-Tóth, Z. Haiman and Z. Frei, *Can quasars, triggered by mergers, account for NANOGrav’s stochastic gravitational wave background?*, *Class. Quant. Grav.* **42** (2025) 075007 [[arXiv:2412.12726](#)] [[INSPIRE](#)].
- [222] V. Rodriguez-Gomez et al., *The merger rate of galaxies in the Illustris simulation: a comparison with observations and semi-empirical models*, *Mon. Not. Roy. Astron. Soc.* **449** (2015) 49 [[arXiv:1502.01339](#)] [[INSPIRE](#)].
- [223] O. Fakhouri, C.-P. Ma and M. Boylan-Kolchin, *The merger rates and mass assembly histories of dark matter haloes in the two millennium simulations*, *Mon. Not. Roy. Astron. Soc.* **406** (2010) 2267 [[arXiv:1001.2304](#)] [[INSPIRE](#)].
- [224] EUCLID collaboration, *Euclid: the $r_b\text{--}M_*$ relation as a function of redshift — I. The $5 \times 109 M_\odot$ black hole in NGC 1272*, *Astron. Astrophys.* **692** (2024) A124 [[arXiv:2411.01927](#)] [[INSPIRE](#)].
- [225] X. Cao and F. Li, *Rapidly spinning massive black holes in active galactic nuclei: evidence from the black hole mass function*, *Mon. Not. Roy. Astron. Soc.* **390** (2008) 561 [[arXiv:0808.0759](#)] [[INSPIRE](#)].
- [226] F. Huang, Y.-C. Bi, Z. Cao and Q.-G. Huang, *Impacts of gravitational-wave background from supermassive black hole binaries on the detection of compact binaries by LISA*, *Chin. Phys. C* **48** (2024) 065103 [[arXiv:2309.14045](#)] [[INSPIRE](#)].
- [227] J. Wolf et al., *First constraints on the AGN X-ray luminosity function at $z \sim 6$ from an eROSITA-detected quasar*, *Astron. Astrophys.* **647** (2021) A5 [[arXiv:2101.05585](#)] [[INSPIRE](#)].
- [228] R. Maiolino et al., *JADES — the diverse population of infant black holes at $4 < z < 11$: merging, tiny, poor, but mighty*, *Astron. Astrophys.* **691** (2024) A145 [[arXiv:2308.01230](#)] [[INSPIRE](#)].