Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 13 November 2007 (MN LATEX style file v2.2)

Modeling the cosmological co-evolution of supermassive black holes and galaxies: I. BH scaling relations and the AGN luminosity function

Federico Marulli[1], Silvia Bonoli[2], Enzo Branchini[3], Lauro Moscardini[1,4], Volker Springel[2]
[1]*Dipartimento di Astronomia, Università degli Studi di Bologna, via Ranzani 1, I-40127 Bologna, Italy*
[2]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Strasse 1, D-85740 Garching, Germany*
[3]*Dipartimento di Fisica, Università degli Studi "Roma Tre", via della Vasca Navale 84, I-00146 Roma, Italy*
[4]*INFN, Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy*

13 November 2007

ABSTRACT
We model the cosmological co-evolution of galaxies and their central supermassive black holes (BHs) within a semi-analytical framework developed on the outputs of the Millennium Simulation. This model, described in detail in Croton et al. (2006) and De Lucia & Blaizot (2007), introduces a 'radio mode' feedback from Active Galactic Nuclei (AGN) at the centre of X-ray emitting atmospheres in galaxy groups and clusters. Thanks to this mechanism, the model can simultaneously explain: (i) the low observed mass drop-out rate in cooling flows; (ii) the exponential cut-off in the bright end of the galaxy luminosity function; and (iii) the bulge-dominated morphologies and old stellar ages of the most massive galaxies in clusters. This paper is the first of a series in which we investigate how well this model can also reproduce the physical properties of BHs and AGN. Here we analyze the scaling relations, the fundamental plane and the mass function of BHs, and compare them with the most recent observational data. Moreover, we extend the semi-analytic model to follow the evolution of the BH mass accretion and its conversion into radiation, and compare the derived AGN bolometric luminosity function with the observed one. While we find for the most part a very good agreement between predicted and observed BH properties, the semi-analytic model underestimates the number density of luminous AGN at high redshifts, independently of the adopted Eddington factor and accretion efficiency. However, an agreement with the observations is possible within the framework of our model, provided it is assumed that the cold gas fraction accreted by BHs at high redshifts is larger than at low redshifts.

1 INTRODUCTION

Over the last years, several observations have demonstrated that supermassive black holes likely reside at the centres of all spheroidal galaxies (see e.g. Kormendy & Richstone 1995; Richstone et al. 1998). Even more interestingly, their properties appear to strongly correlate with those of their host galaxies (Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Graham et al. 2001; Tremaine et al. 2002; McLure & Dunlop 2002; Baes et al. 2003; Marconi et al. 2004; Häring & Rix 2004; Feoli & Mele 2007; Graham & Driver 2007) and, apparently, also with the ones of the whole host dark matter (DM) haloes (Ferrarese 2002; Baes et al. 2003; Ferrarese & Ford 2005). Although it is not yet clear which of these relations is "more fundamental" (see e.g. Novak et al. 2006), they reasonably suggest a close link between the assembly history of the BHs and the cosmological evolution of galaxies. Most recently, Hopkins et al. (2007a) have shown that these relationships are not independent and could be interpreted as different projections of a BH fundamental plane, analogous to the fundamental plane of elliptical galaxies. The striking similarity between these two fundamental planes is another clue that galaxy spheroids and BHs do not form and evolve independently.

The paradigm that AGN are powered by mass accretion onto BHs (Salpeter 1964; Lynden-Bell 1969) has also received very strong support from spectroscopic and photometric observations of the stellar and gas dynamics in the central regions of local spheroidal galaxies and bulges. Moreover, by estimating the total energy radiated by AGN during their whole life, it can be shown that nearly all the mass in BHs has been accumulated during periods of bright AGN activity (Soltan 1982), implying that the common physical process which produces galaxy spheroids and BHs must also be responsible for triggering bright AGN.

Such a cosmological co-evolution of BHs, AGN and galaxies is expected in the standard framework, in which cosmic structures grow hierarchically via gravitational instability and merging events destabilize the gas at the galaxy centres, triggering star formation and BH mass accretion. In order to investigate this complex scenario, several models have been developed, based on either pure analytic approximations (see, e.g., Efstathiou & Rees 1988; Haehnelt & Rees 1993; Haiman & Loeb 1998; Percival & Miller 1999; Haiman & Menou

2000; Martini & Weinberg 2001; Wyithe & Loeb 2003; Hatziminaoglou et al. 2003; Hopkins et al. 2007b), or semi-analytic ones (see, e.g., Kauffmann & Haehnelt 2000; Cavaliere & Vittorini 2002; Enoki et al. 2003; Volonteri et al. 2003; Granato et al. 2004; Springel et al. 2005; Cattaneo et al. 2005; Croton et al. 2006; Malbon et al. 2006; Fontanot et al. 2006). Recently, thanks to the availability of unprecedented computational power, fully numerical models have also become available (see, e.g., Di Matteo et al. 2005; Springel et al. 2005; Li et al. 2007; Sijacki et al. 2007; Di Matteo et al. 2007).

Simple analytic models in which AGN activity is only triggered by DM halo major mergers succeeded in quantitatively describing the observed evolution of the AGN number counts and luminosity at all but low redshifts, provided that some mechanism is advocated to inhibit accretion within massive haloes hosting bright AGN. However they fail in reproducing the observed AGN clustering at high redshifts (Marulli et al. 2006). Slightly more sophisticated semi-analytic models in which the halo merger history and associated BHs are followed by Monte Carlo realizations of the merger hierarchy, while the baryonic physics is neglected as well, can correctly reproduce both the AGN luminosity and clustering function at $z \gtrsim 1$ (Marulli et al. 2006), but the number density of faint AGN is significantly below observations, a clear indication that DM halo mergers cannot constitute the only trigger to accretion episodes in the local BH population (Marulli et al. 2007), and that in order to properly describe the cosmological evolution of BHs and AGN, the main baryonic phenomena involving the gas contents of DM halos cannot be neglected.

This complication is reminiscent of the one found in the description of galaxies, where the well-known mismatch in shape between the predicted distribution of DM halo masses and the observed distribution of galaxy luminosities requires the consideration of complex baryonic phenomena like, for instance, cooling inefficiencies to reduce gas condensation in massive structures, or supernova (White & Rees 1978; White & Frenk 1991) and stellar kinetic feedback (Fontanot et al. 2006) to remove cold gas in low mass systems, as well as photoionisation heating to suppress the formation of dwarfs (Efstathiou 1992). Cooling effects alone are however too weak to produce the bright end cut-off of the luminosity function, and it seems to be mandatory to include additional feedback processes in massive halos (e.g. Benson et al. 2003; Fontanot et al. 2006; Croton et al. 2006; Ciotti & Ostriker 2007). Standard models of galaxy formation face two additional problems: i) the persistence of a hot gas atmosphere at the centre of most galaxy clusters despite the fact that the local cooling time is much shorter than the age of the system (see, e.g. Cowie & Binney 1977; Fabian & Nulsen 1977; Peterson et al. 2001; Tamura et al. 2001; Fabian et al. 2003; McNamara et al. 2005; Morandi & Ettori 2007, and references therein), and ii) the fact that most massive galaxies, typically ellipticals in clusters, are made of the oldest stars and so finished their star formation earlier than lower mass galaxies (see, e.g. Cowie et al. 1996; De Lucia et al. 2006; Cimatti et al. 2006, and references therein).

In this paper we study the cosmological co-evolution of galaxies and their central BHs using a semi-analytical model developed on the outputs of the Millennium Simulation and described in detail in Croton et al. (2006) and De Lucia & Blaizot (2007). In this scenario, *radio mode* feedback from AGN at the centre of galaxy groups and clusters is invoked to prevent significant gas cooling in large halos, thus limiting the mass of the central galaxies and preventing them from forming stars at late times when their mass and morphology can still change through mergers. Thanks to this mechanism, Croton et al. (2006) demonstrated that such a model can simultaneously explain the low observed mass drop-out rate in cooling flows, the exponential cut-off in the bright-end of the galaxy luminosity function, and the bulge-dominated morphologies and stellar ages of the most massive galaxies in clusters.

Here we are interested in investigating how well this model can also reproduce the statistical properties of BHs and AGN. To do that, we extend the original model by adding new semi-analytical prescriptions to describe the BH mass accretion rate in the accretion episodes triggered by galaxy mergers, which fuel the *quasar mode*, and their conversion into radiation. We then analyze the scaling relations, the fundamental plane and the mass function of BHs, and compare them with the most recent observational data available. Finally, we compare the predicted AGN bolometric luminosity function with the observed one, and propose some modifications to the original semi-analytic assumptions to better fit the data.

The paper is organized as follows. In Section 2, we briefly describe the main aspects of our semi-analytic model and illustrate the new equation introduced to describe the BH mass accretion in the *quasar mode* in more detail. In Section 3, we compare the model predictions with the best observational data available for the BH and AGN populations. Finally, in Section 4 we summarize our conclusions.

2 THE MODEL

Our semi-analytic model for the co-evolution of DM haloes, galaxies and their central BHs consists of three ingredients, that we describe separately in this section: a numerical simulation to obtain the merger history of the DM haloes, a set of analytic prescriptions to trace the evolution of galaxies within their host haloes and a set of recipes to follow the BH accretion history and the AGN phenomenon.

2.1 Numerical simulation

In this work we use the outputs of the Millennium Simulation, which followed the dynamical evolution of $2160^3 \simeq 10^{10}$ DM particles with mass $8.6 \times 10^8\, h^{-1}\mathrm{M}_\odot$ in a periodic box of $500\, h^{-1}\mathrm{Mpc}$ on a side, in a ΛCDM "concordance" cosmological framework (Springel et al. 2005). The computational box is large enough to include rare objects such as quasars or rich galaxy clusters, the largest of which contain about 3 million simulation particles at $z = 0$. At the same time, the mass resolution is high enough to resolve the DM halo of $0.1 L_\star$ galaxies with ~ 100 particles. The short-range gravitational force law is softened on the co-moving scale $5 h^{-1}\mathrm{kpc}$ (Plummer-equivalent) which may be taken as the spatial resolution limit of the calculation. The cosmological parameters (the matter density parameter $\Omega_\mathrm{m} = 0.25$, the baryon density parameter $\Omega_\mathrm{b} = 0.045$, the Hubble parameter $h = H_0/100\,\mathrm{km\,s^{-1}Mpc^{-1}} = 0.73$, the cosmological constant contribution to the density parameter $\Omega_\Lambda = 0.75$, the primordial spectral index $n = 1$, and the power spectrum normalization $\sigma_8 = 0.9$), are consistent with determinations from the combined analysis of the 2-degree Field Galaxy Redshift Survey (2dFGRS) (Colless et al. 2001) and first-year WMAP data (Spergel et al. 2003), as shown by Sánchez et al. (2006). We recall that the more recent analysis of the WMAP 3-year data

(Spergel et al. 2007) suggests slightly different values (in particular smaller values for $\Omega_{\rm m}$, σ_8 and n). However, as demonstrated by Wang et al. (2007), due to the current modelling uncertainties, it is not possible to distinguish the two WMAP cosmologies on the basis of the observed galaxy properties, since the variations induced by acceptable modifications of the free parameters of the galaxy formation model are at least as large as those produced by the variation in the cosmological parameters.

The Millennium Simulation was carried out with a special version of the GADGET-2 code (Springel 2005), optimized for very low memory consumption, at the Computing Centre of the Max-Planck Society in Garching, Germany. We make use of hierarchical merging trees extracted from this simulation which encode the full formation history of DM haloes and subhalos, previously identified with, respectively, a friends-of-friends (FOF) group-finder and an extended version of the SUBFIND algorithm (Springel et al. 2001). These trees constitute the backbone of our semi-analytic model, which is implemented during the post-processing phase: this allows us to simulate the wide range of baryonic processes occurring during the formation and evolution of galaxies and their central BHs.

2.2 Galaxy evolution

We use the galaxy formation model of Croton et al. (2006) as updated by De Lucia & Blaizot (2007). Although not in agreement with some properties of the red and blue galaxy populations (see, e.g., Weinmann et al. 2006; Kitzbichler & White 2007), this model is able to reproduce the overall observed properties of galaxies, i.e. the relations between stellar mass, gas mass and metallicity, the luminosity, colour and morphology distributions (Croton et al. 2006; De Lucia et al. 2006), the two-point galaxy correlation functions (Springel et al. 2005), and the global galaxy luminosity and mass functions at high redshift (Kitzbichler & White 2007). We refer to the original papers for a full description of the numerical implementation of the model. In the following, we briefly recall the treatment of the physical processes involved in the galaxy evolution, and describe the prescriptions for the BH growth and the AGN evolution.

Following the standard paradigm set out by White & Frenk (1991) and adapted to high-resolution N-body simulations by Springel et al. (2001), we assume that as a DM halo collapses, a fraction $f_b = 0.17$ of its mass is in the form of baryons and collapses with it, consistent with the first-year WMAP result (Spergel et al. 2003). Initially, these baryons are in the form of a diffuse gas with primordial composition, but later they include gas in several phases as well as stars and heavy elements. Conventionally, with the simplifying assumption of an ideal gas which cools isobarically, the cooling time of the gas is computed as the ratio of its specific thermal energy to the cooling rate per unit volume,

$$t_{\rm cool} = \frac{3}{2}\frac{\bar{\mu} m_p k T}{\rho_g(r)\Lambda(T,Z)} \ , \qquad (1)$$

where $\bar{\mu} m_p$ is the mean particle mass, k is the Boltzmann constant, $\rho_g(r)$ is the hot gas density, and $\Lambda(T,Z)$ is the cooling function (Sutherland & Dopita 1993; Maio et al. 2007). Equation (1) is valid at temperature higher than $\sim 10^4$ K, where hydrogen and helium remain ionized and the number of particles remains approximately constant.

We assume the post-shock temperature of the infalling gas to be the virial temperature of the halo, $T = 35.9\,(V_{\rm vir}/{\rm km\,s^{-1}})^2$ K, where $V_{\rm vir}$ is the halo virial velocity. Moreover, we assume that the hot gas within a static atmosphere has a simple 'isothermal' distribution,

$$\rho_g(r) = \frac{m_{\rm hot}}{4\pi R_{\rm vir} r^2} \ , \qquad (2)$$

where $m_{\rm hot}$ is the total hot gas mass associated with the halo and is assumed to extend to its virial radius $R_{\rm vir}$.

In order to estimate an instantaneous cooling rate onto the central object of a halo, given its current hot gas content, we define the cooling radius, $r_{\rm cool}$, as the radius at which the local cooling time (assuming the structure of equation (2)) is equal to the halo dynamical time, $R_{\rm vir}/V_{\rm vir} = 0.1 H(z)^{-1}$ (Springel et al. 2001; De Lucia et al. 2004; Croton et al. 2006); here $H(z)$ represents the redshift evolution of the Hubble constant. The cooling rate can then be determined through the following continuity equation,

$$\dot{m}_{\rm cool} = 4\pi\rho_g(r_{\rm cool}) r_{\rm cool}^2 \dot{r}_{\rm cool} \ . \qquad (3)$$

More details about our cooling prescriptions can be found in Croton et al. (2006).

The photo-ionization heating of the intergalactic medium suppresses the concentration of baryons in shallow potentials (Efstathiou 1992), and can be responsible of the inefficient accretion and cooling in low-mass haloes. Following Gnedin (2000), we model the effect of such photo-ionization heating by defining a characteristic mass scale, $M_{\rm F}$, below which the gas fraction f_b is reduced relatively to the universal value:

$$f_{\rm b}^{\rm halo}(z, M_{\rm vir}) = \frac{f_{\rm b}^{\rm cosmic}}{[1+0.26 M_{\rm F}(z)/M_{\rm vir}]^3} \ . \qquad (4)$$

We adopt the $M_{\rm F}(z)$ parameterization of Kravtsov et al. (2004), which results in a filtering mass $M_{\rm F}$ of $4\times10^9 M_\odot$ at the reionization epoch, and $3\times10^{10} M_\odot$ by the present day (but see Hoeft et al. 2006).

In the semi-analytic framework we use in this work, the star formation is assumed to occur at a rate given by:

$$\dot{m}_* = \alpha_{\rm SF}(m_{\rm cold} - m_{\rm crit})/t_{\rm dyn,disc} \ , \qquad (5)$$

where $m_{\rm cold}$ is the cold gas mass, $t_{\rm dyn,disc}$ is the dynamical time of the galaxy, defined as the ratio between the disk radius and the virial velocity, $m_{\rm crit}$ corresponds to a critical value for the gas surface density (Kauffmann 1996; Kennicutt 1998; Mo et al. 1998), and $\alpha_{\rm SF} = 0.03$ controls the efficiency of the transformation of cold gas into stars. Massive stars explode as supernovae shortly after star formation events and are assumed to reheat a gas mass proportional to the mass of stars:

$$\Delta m_{\rm reheated} = \varepsilon_{\rm disk} \Delta m_* , \qquad (6)$$

where we set the free parameter $\varepsilon_{\rm disk} = 3.5$ based on the observational data. The energy released by an event which forms a mass Δm_* in stars is assumed to be:

$$\Delta E_{\rm SN} = 0.5 \varepsilon_{\rm halo} \Delta m_* V_{SN}^2 , \qquad (7)$$

where $0.5 V_{\rm SN}^2$ is the mean supernova energy injected per unit mass of newly formed stars, and $\varepsilon_{\rm halo}$ represents the efficiency with which this energy is able to convert cold interstellar medium into hot, diffuse halo gas. The amount of gas that leaves the DM halo in a "super-wind" is determined by computing whether excess SN energy is available to drive the flow after reheating of material to the halo virial temperature.

We model the disk instabilities using the analytic stability criterion of Mo et al. (1998); the stellar disk of a galaxy becomes unstable when the following inequality is met:

$$\frac{V_c}{(G m_{\rm disk}/r_{\rm disk})^{1/2}} \leqslant 1 \ . \qquad (8)$$

At each time-step we evaluate the left-hand side of equation (8) for each galaxy, and if it is smaller than unity we transfer enough stellar mass from disk to bulge (at fixed $r_{\rm D}$) to restore stability.

In the Millennium Run, substructures are followed down to masses of $1.7\times10^{10}h^{-1}M_{\odot}$, so that we can properly follow the motion of galaxies inside their hosting DM haloes until tidal truncation and stripping disrupt their subhalos at this resolution limit. At this point, we estimate a survival time for the galaxies using their current orbit and the dynamical friction formula of Binney & Tremaine (1987) multiplied by a factor of 2, as in De Lucia & Blaizot (2007). After this time, the galaxy is assumed to merge onto the central galaxy of its own halo. Galaxy mergers induce starburst which we describe using the "collisional starburst" prescription introduced by Somerville et al. (2001). In this model, a fraction $e_{\rm burst}$ of the combined cold gas from the two merging galaxies is turned into stars as follows:

$$e_{\rm burst}=\beta_{\rm burst}(m_{\rm sat}/m_{\rm central})^{\alpha_{\rm burst}}\,, \qquad (9)$$

where the two parameters are taken as $\alpha_{\rm burst}=0.7$ and $\beta_{\rm burst}=0.56$, appropriate for merger mass ratios ranging from 1:10 to 1:1 (Cox 2004).

2.3 BH mass accretion and AGN

2.3.1 The 'radio mode'

When a static hot halo has formed around a galaxy, we assume that a fraction of the hot gas continuously accretes onto the central BH, causing a low-energy 'radio' activity in the galaxy centre. Following Croton et al. (2006), the BH mass accretion rate during these phases is postulated to scale as follows:

$$\dot{M}_{\rm BH,R}=\kappa_{\rm AGN}\left(\frac{M_{\rm BH}}{10^8M_{\odot}}\right)\left(\frac{f_{\rm hot}}{0.1}\right)\left(\frac{V_{\rm vir}}{200\,{\rm km\,s^{-1}}}\right)^3\,, \qquad (10)$$

where $M_{\rm BH}$ is the BH mass, $f_{\rm hot}$ is the fraction of the total halo mass in the form of hot gas, and $\kappa_{\rm AGN}$ is a free parameter set equal to $7.5\times10^{-6}{\rm M_{\odot}yr^{-1}}$ in order to reproduce the turnover at the bright end of the galaxy luminosity function. Since $f_{\rm hot}$ is approximately constant for $V_{\rm vir}\gtrsim150\,{\rm km\,s^{-1}}$, the dependence of $\dot{m}_{\rm BH,R}$ on this quantity has a little effect. Note that the accretion rate given by equation (10) is typically orders-of-magnitude below the Eddington limit. In fact, the total mass growth of BHs in the radio relative to the quasar mode discussed below is negligible.

It is also assumed that the *radio mode* feedback injects energy efficiently into the surrounding medium, which can reduce or even stop the cooling flow in the halo centres. The mechanical heating generated by this kind of BH mass accretion and described as $L_{\rm BH}=\varepsilon\dot{M}_{\rm BH}c^2$, where $\varepsilon=0.1$ is the *accretion efficiency* and c is the speed of light, induces a modified infall rate of the following kind:

$$\dot{m}'_{\rm cool}=\dot{m}_{\rm cool}-\frac{L_{\rm BH}}{0.5V_{\rm vir}^2}\,. \qquad (11)$$

For consistency we never allow $\dot{m}'_{\rm cool}$ to fall below zero. In this scenario, the effectiveness of radio AGN in suppressing cooling flows is greatest at late times and for large values of the BH mass, which is required to successfully reproduce the luminosities, colours and clustering of low-redshift bright galaxies.

2.3.2 The 'quasar mode'

In our model BHs accrete mass after a galaxy merger both through coalescence with another BH and by accreting cold gas, the latter being the dominant accretion mechanism. For simplicity, the BH coalescence is modelled as a direct sum of the progenitor masses, thus ignoring gravitational wave losses. Following Kauffmann & Haehnelt (2000), we assume that the gas mass accreted during a merger is proportional to the total cold gas mass present, but with an efficiency which is lower for smaller mass systems and for unequal mergers:

$$\Delta M_{\rm BH,Q}=\frac{f'_{\rm BH}\,m_{\rm cold}}{1+(280\,{\rm km\,s^{-1}}/{\rm V_{vir}})^2}\,, \qquad (12)$$

where

$$f'_{\rm BH}=f_{\rm BH}\,(m_{\rm sat}/m_{\rm central})\,, \qquad (13)$$

and $f_{\rm BH}\approx0.03$ is chosen to reproduce the observed local $M_{\rm BH}-M_{\rm bulge}$ relation. Thus, any merger-induced perturbation to the gas disk (which might come from a bar instability or a merger-induced starburst) can in principle drive gas onto the central BH. However, the fractional contribution of minor mergers is typically quite small, so that accretion driven by major mergers is the dominant mode of BH growth in our scenario. This kind of accretion, which we call *quasar mode*, is also closely associated with starbursts, which occur concurrently. We do not model feedback from the quasar activity in the current model, but it can be approximately represented by an enhanced effective feedback efficiency for the supernovae associated with the intense starburst.

2.3.3 AGN luminosity

The output of the model summarized hitherto, called *DeLucia2006a catalogue* (De Lucia & Blaizot 2007), is publicly available at http://www.mpa-garching.mpg.de/millennium (Lemson & Virgo Consortium 2006). In this default model, for simplicity, the BH mass accretion triggered by each merger is implemented as an instantaneous event and the BH seed masses are set equal to zero.

In order to study the evolution of AGN inside this cosmological framework, we improve the original model of De Lucia & Blaizot (2007) by adding new semi-analytical prescriptions to describe the BH mass accretion rate during each merger event in the *quasar mode*, and its conversion into radiation. In this implementation, BHs do not accrete mass instantaneously. Instead, the accretion is coupled to the light curve model adopted. If a galaxy undergoes a merger while the central BH is still accreting mass from a previous merger, the cold gas still to be accreted is added to the new gas reservoir, and the accretion re-starts under the new physical conditions. In Sect. 3.1.4 we show that the BH scaling relations are weakly affected by this change. We use the following definitions to parameterize the bolometric luminosity emitted by accretion onto BHs, as a function of the *accretion efficiency*, ε, and the *Eddington factor*, $f_{\rm Edd}(t):=L_{\rm bol}(t)/L_{\rm Edd}(t)$,

$$\begin{aligned} L_{\rm bol}(t) &= \frac{\varepsilon}{1-\varepsilon}\dot{M}_{\rm BH}(t)c^2 \\ &= f_{\rm Edd}(t)L_{\rm Edd}(t)=f_{\rm Edd}(t)\frac{M_{\rm BH}(t)}{t_{\rm Edd}}c^2, \\ &\Longrightarrow d\ln M_{\rm BH}(t)=\frac{dt}{t_{\rm ef}(t)}, \end{aligned} \qquad (14)$$

where $L_{\rm Edd}$ is the Eddington luminosity, $t_{\rm Edd}=\sigma_T c/(4\pi m_p G)\sim0.45\,{\rm Gyr}$ and $t_{\rm ef}(t)=\frac{\varepsilon}{1-\varepsilon}\frac{t_{\rm Edd}}{f_{\rm Edd}(t)}$ is the e-folding time ($t_{\rm ef}\equiv t_{\rm Salpeter}$ if $f_{\rm Edd}=1$).

No strong observational constraints are available for ε and

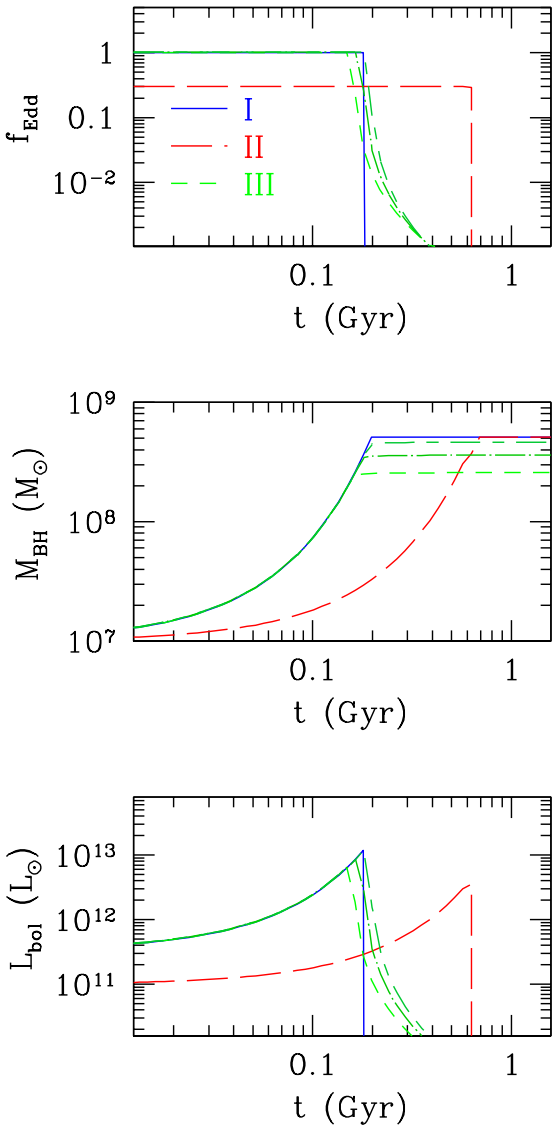

Figure 1. The time evolution of $f_{\rm Edd}$ (top panel), $M_{\rm BH}$ (central panel) and $L_{\rm bol}$ (bottom panel) for our three lightcurve models (*I* (blue solid lines), *II* (red short-dashed lines) and *III* (green lines)), for an illustrative case of a BH of mass $M_{\rm BH}=10^7 M_\odot$ accreting a mass $\Delta M_{\rm BH,Q}=5\times10^8 M_\odot$, starting at $z=3$. The three green curves, showing our model *III*, have been obtained by setting $\mathcal{F}=0.5$ (short dashed), 0.7 (dotted-long dashed) and 0.9 (short dashed-long dashed).

$f_{\rm Edd}$, the parameters that regulate the BHs powering the AGN and, more importantly, if and how they depend on redshift, BH masses, AGN luminosities and so on. However, some observations at $z\sim0$ indicate that $0.04<\epsilon<0.16$ and $0.1<f_{\rm Edd}<1.7$ (Marconi et al. 2004). Furthermore, it has been suggested that $f_{\rm Edd}$ may depends on redshift (Shankar et al. 2004) and BH mass (Netzer & Trakhtenbrot 2007). In this paper, for simplicity, we do not follow the evolution of the BH spins (see, e.g. Volonteri et al. 2007, and references therein) and we take a constant mean value for the accretion efficiency of $\epsilon=\langle\epsilon\rangle=0.1$ at all redshifts.

For $f_{\rm Edd}$, which determines the lightcurves associated with individual quasar events, we consider instead three different prescriptions:

- *I*: $f_{\rm Edd}=1$, the simplest possible assumption. Here the quasar is either 'on' at its maximum Eddington luminosity, or 'off'.
- *II*:

$$f_{\rm Edd}(z)=\begin{cases} f_{\rm Edd,0} & z\geqslant 3 \\ f_{\rm Edd,0}\cdot[(1+z)/4]^{1.4} & z<3 \end{cases} \qquad (15)$$

with $f_{\rm Edd,0}=0.3$, as suggested by Shankar et al. (2004) to match the BH mass function derived from a deconvolution of the AGN luminosity function and the local BH mass function.

- *III*: based on the analysis of self-consistent hydrodynamical simulations of galaxy mergers, Hopkins et al. (2005) noticed that the light curves of active BHs are complex, showing periods of rapid accretion after "first passage" of the merging galaxies, followed by a long-lasting quiescent phase, then a transition to a highly luminous, peaked quasar phase, finally a fading away when quasar feedback expels gas from the remnant's centre in a self-regulated mechanism after the BH reaches a critical mass. In spite of this complexity, as a first order approximation, the typical evolution of an active BH can be simply described as a two-stage process of a rapid, Eddington-limited growth up to a peak BH mass, preceeded and followed by a much longer quiescent phase with lower Eddington ratios. In this latter phase, the average time spent by AGN per logarithmic luminosity interval can be approximated as (Hopkins et al. 2005)

$$\frac{{\rm d}t}{{\rm d}\ln L_{\rm bol}}=|\alpha|t_9\left(\frac{L_{\rm bol}(t)}{10^9L_\odot}\right)^\alpha, \qquad (16)$$

where $t_9\equiv t_Q(L'>10^9L_\odot)$ and $t_Q(L'>L)$ is the total AGN lifetime above a given luminosity L; $t_9\sim10^9$ yr over the range $10^9L_\odot<L_{\rm bol}<L_{\rm peak}$. In the range $10^{10}L_\odot\lesssim L_{\rm peak}\lesssim10^{14}L_\odot$, Hopkins et al. (2005) found that α is a function of only the AGN luminosity at the peak of its activity, $L_{\rm peak}$, given by $\alpha=-0.95+0.32\log(L_{\rm peak}/10^{12}L_\odot)$, with $\alpha=-0.2$ (the approximate slope of the Eddington-limited case) as an upper limit. We here interpret the Hopkins model as describing primarily the decline phase of the quasar activity, after the black hole has grown at the Eddington rate to a peak mass $M_{\rm BH,peak}=M_{\rm BH}(t_{\rm in})+\mathcal{F}\cdot\Delta M_{\rm BH,Q}\cdot(1-\epsilon)$, where $M_{\rm BH}(t_{\rm in})$ is the initial BH mass and $\Delta M_{\rm BH,Q}$ is the fraction of cold gas mass accreted. Here $\mathcal{F}$ is an additional free parameter, in the range $0\leqslant\mathcal{F}\leqslant1$. For $\mathcal{F}=1$ the BH emits at the Eddington rate. In the opposite limit ($\mathcal{F}=0$) the AGN reaches instantaneously a peak luminosity, and the whole light curve is described by equation (16). We found that $\mathcal{F}=0.7$ is the value that best matches the AGN luminosity function. We note that this interpretation of the Hopkins model is plausible but not unique, as part of the time described by equation (16) could also be associated with the rising part of the lightcurve.

From equation (16) and with the following definition

$$\tilde{f}_{\rm Edd}(t):=\frac{L_{\rm bol}(t)}{L_{\rm peak}}=f_{\rm Edd}(t)\frac{L_{\rm Edd}(t)}{L_{\rm peak}}, \qquad (17)$$

we can derive:

$$\frac{{\rm d}\tilde{f}_{\rm Edd}(t)}{{\rm d}t}=-\frac{\tilde{f}_{\rm Edd}^{1-\alpha}(t)}{\alpha t_9}\left(\frac{L_{\rm peak}}{10^9L_\odot}\right)^{-\alpha}, \qquad (18)$$

$$\Longrightarrow\tilde{f}_{\rm Edd}(t)=\left[\tilde{f}_{\rm Edd,0}^{\alpha}+\left(\frac{L_{\rm peak}}{10^9L_\odot}\right)^{-\alpha}\frac{t}{t_9}\right]^{1/\alpha}. \qquad (19)$$

Here we neglected the absolute value of α present in equation (16),

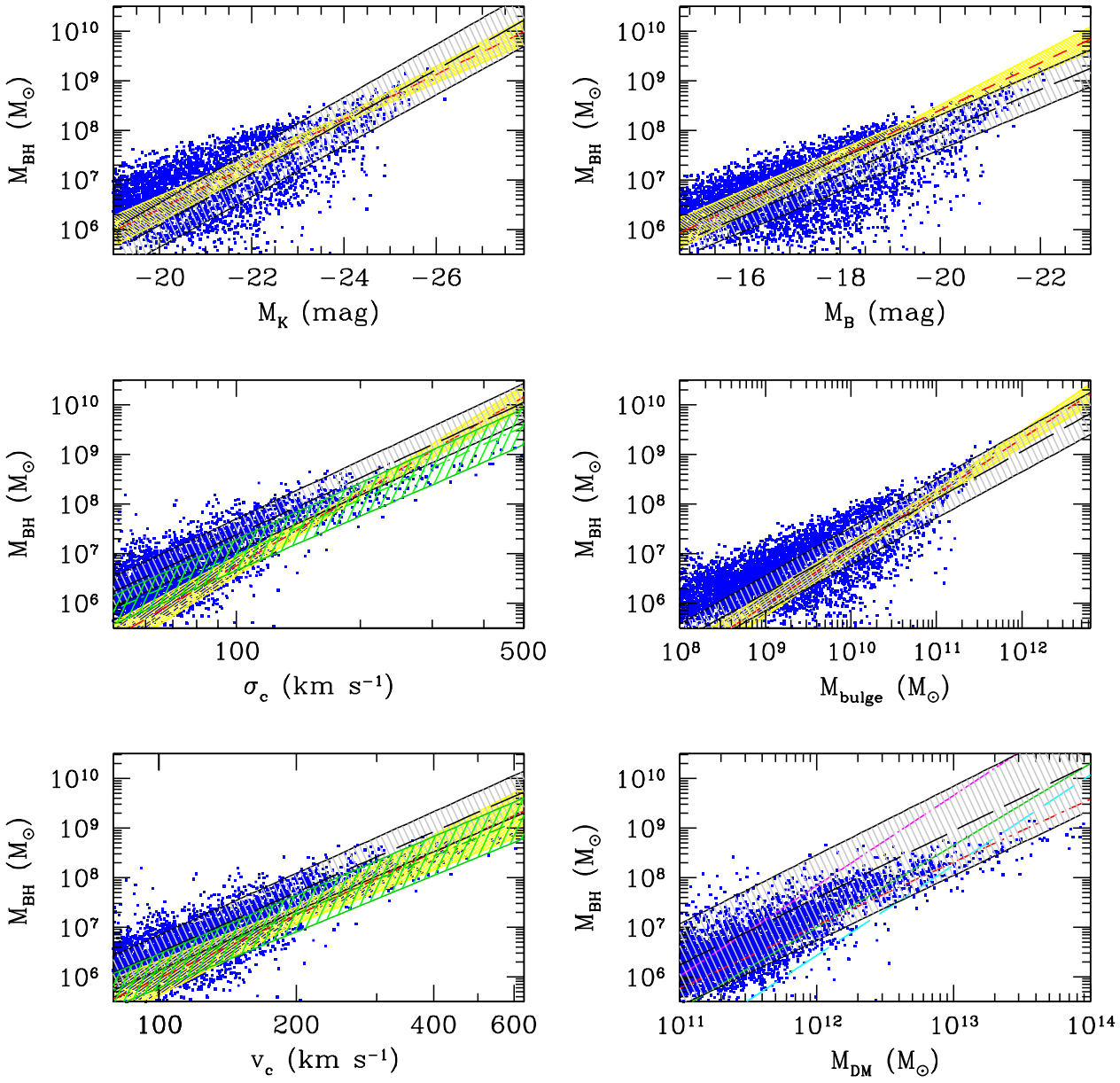

Figure 2. Starting from the upper left panel down to the bottom right one, scaling relations between the masses of the central BHs in the simulated galaxies with six different properties of their hosts: the K- and B-band bulge magnitude (top left and right panels, respectively), the bulge velocity dispersion and mass (central left and right panels, respectively), the circular velocity of the galaxy (bottom left panel) and the virial mass of the DM halo (bottom right panel). Blue dots represent the outputs of the *DeLucia2006a catalogue*, grey and yellow shaded areas show the best fit to the model predictions and to the observational datasets, respectively. Starting from the upper left panel down to the lower right, the yellow shaded areas refer to the best-fit relations obtained by Marconi et al. (2004) (the upper two panels of the plot), Ferrarese & Ford (2005), Häring & Rix (2004), Baes et al. (2003) and, in the lower-right panel, the four curves show the equations 4 (cyan), 6 (green) and 7 (magenta) of Ferrarese (2002) and the results of Baes et al. (2003) (red).

for the purpose of having $\tilde{f}_{\rm Edd}(t)$ a decreasing function of time. Finally, from equations (14), (17) and (19), we have:

$$M_{\rm BH}(t) = M_{\rm BH,peak} + \frac{A}{BC}\left[(1+Ct)^B - 1\right], \qquad (20)$$

where $A = \frac{1-\varepsilon}{\varepsilon}\frac{M_{\rm BH,peak}}{t_{\rm Edd}}$, $B = \frac{1}{\alpha}+1$, $C = \left(\frac{L_{\rm peak}}{10^9 L_\odot}\right)^{-\alpha}\frac{1}{t_9}$. To derive equation (20) we set $\tilde{f}_{\rm Edd,0} = 1$ for continuity. We also impose $f_{\rm Edd} = 10^{-3}$ as lower limit for the Eddington factor.

Figure 1 shows the evolution of $f_{\rm Edd}(t)$ (top panel), $M_{\rm BH}(t)$ (central panel) and $L_{\rm bol}(t)$ (bottom panel) for an illustrative case of a BH of $M_{\rm BH} = 10^7 M_\odot$ accreting a mass $M_{\rm accr} = 5\times10^8 M_\odot$, starting at $z = 3$, in the three prescriptions considered. The three

Relation	Normalization (α)	Slope (β)	Scatter	$\text{Scatter}_{\text{corrected}}$
$\log(M_{\rm BH}) - M_{\rm K}$	-4.37(0.24)	-0.52(0.01)	0.68	0.53
$\log(M_{\rm BH}) - M_{\rm B}$	-0.61(0.17)	-0.43(0.01)	0.62	0.53
$\log(M_{\rm BH}) - \log(\sigma_{\rm c})$	-0.26(0.16)	3.82(0.08)	0.42	0.28
$\log(M_{\rm BH}) - \log(M_{\rm bulge})$	-2.39(0.19)	0.96(0.02)	0.58	0.50
$\log(M_{\rm BH}) - \log(V_{\rm c})$	-1.61(0.18)	4.05(0.09)	0.45	
$\log(M_{\rm BH}) - \log(M_{\rm DM})$	-8.61(0.42)	1.35(0.04)	0.50	

Table 1. Parameters of the linear fits to the scaling relations shown in Figure 2. A correlation of the form $y = \alpha + \beta \cdot x$ has been assumed for all relations. The uncertainties in the normalizations and in the slopes are shown in parentheses. For details about the computation of the Scatter and the $\text{Scatter}_{\text{corrected}}$ see Sect. 3.1.

Relation	α	β	γ	Scatter
$\log(M_{\rm BH}) - M_{\rm K}$	17.29(0.10)	1.25(0.01)	0.04(0.01)	0.51
$\log(M_{\rm BH}) - M_{\rm K}$	9.81(0.03)	0.63(0.01)	0.03(0.01)	0.47
$\log(M_{\rm BH}) - \log(M_{\rm bulge})$	14.16(0.07)	-2.21(0.01)	0.15(0.01)	0.44

Table 2. Parameters of the fits to the scaling relations shown in Figure 3. A correlation of the form $y = \alpha + \beta \cdot x + \gamma \cdot x^2$ has been assumed for all three relations. The uncertainties in the parameters are shown in parentheses. For details about the computation of the Scatter see Sect. 3.1.

green curves refer to lightcurve model *III*, in which we set $\mathcal{F} = 0.5$ (short dashed), $= 0.7$ (dot-long dashed) and $= 0.9$ (short dashed-long dashed).

Due to the present uncertainties concerning the origin of the BH seeds and their mass distribution, we assume $M_{\rm BH,seed} = 10^3 M_{\odot}$ for all seed BHs, irrespective of their halo host properties and their origin. Our results are robust with respect to this hypothesis since, as we have verified, they are basically unaffected by varying $M_{\rm BH,seed}$ in the range $[10^2 - 10^5] M_{\odot}$ at $z \lesssim 3$. More significant differences occur at higher redshifts, which we will investigate in detail in future work.

The main parameters of our model are listed in Table 1 of Croton et al. (2006), with the exception of, as in De Lucia & Blaizot (2007), the values for the quiescent hot gas BH accretion rate, $\kappa_{\rm AGN}$ (defined in section 2.3.1), the star formation efficiency $\alpha_{\rm SF}$ of equation (5), and the instantaneous recycled fraction of star formation to the cold disk, R, which we set equal to 0.43 (see Section 3.9 of Croton et al. 2006).

3 MODELS VS. OBSERVATIONS

3.1 The BH scaling relations

Several observational evidences indicate that the masses of the BHs hosted at the centres of galaxies strongly correlate with different properties of their host bulges and DM haloes. In this section we compare the most recently observed BH scaling relations at $z = 0$ with the predictions of the original model of De Lucia & Blaizot (2007), i.e. the predictions we obtain when assuming instantaneous mass accretion. We explore the effect of specifying the mass accretion rate at the end of this section.

3.1.1 One-parameter relations

In Figure 2, we show the correlation between the masses of the model BHs with six properties of their hosts, the K- and B-band bulge magnitude ($\rm M_B$ and $\rm M_K$), the bulge mass and velocity dispersion ($\rm M_{bulge}$ and $\sigma_{\rm c}$), the circular velocity of the galaxy and the virial mass of the DM halo ($\rm V_c$ and $\rm M_{DM}$). The blue dots represent the outputs of the model, while grey and yellow shaded areas show linear best fits to the model predictions and to the observational datasets, respectively.

The dots in the plot refer to the population of BHs hosted in the central galaxies of FOF groups, or subhalos. We do not include those in satellite galaxies since in this case the host properties cannot be determined accurately. The data we have considered are: the $\rm M_{BH} - M_B$ and $\rm M_{BH} - M_K$ relations of Marconi et al. (2004) (top panels) the $\rm M_{BH} - \sigma_c$ relation of Ferrarese & Ford (2005) (central left) the $\rm M_{BH} - M_{bulge}$ relation of Häring & Rix (2004) (central right) and the $\rm M_{BH} - V_c$ relation of Baes et al. (2003) (bottom left). No direct observational estimate is available for the $\rm M_{BH} - M_{DM}$ relation shown in the bottom right panel. The curves shown in this panel have been derived using different assumptions for the $\rm M_{DM} - V_c$ relation. In particular, the cyan, green and magenta lines correspond to equations (4), (6) and (7) of Ferrarese (2002), while the red curve is taken from Baes et al. (2003).

Model predictions for $\rm V_c$ and $\sigma_{\rm c}$ have been obtained by adopting two different assumptions: i) $\rm V_c = V_{max}$, where $\rm V_{max}$ is the maximum rotational velocity of the subhalo hosting the galaxy at its centre, and ii) $\rm V_c = 1.8\, V_{vir}$ as derived by Seljak (2002). The bulge velocity dispersion $\sigma_{\rm c}$ is derived from the $\rm V_c - \sigma_c$ relation of Baes et al. (2003). In the bottom panels, the grey areas correspond to a circular velocity obtained through hypothesis i) while the green ones, in better agreement with the data, assume hypothesis ii).

The linear fit to the model data has been obtained using the bisector modification to the ordinary least squares minimization approach, proposed by Akritas & Bershady (1996), for which the best-fit results correspond to the bisection of those obtained from minimizations in the vertical and horizontal directions. The estimator is robust and has the advantage of taking into account the possible intrinsic scatter in the relation. The values of the best fit slope and the normalization are listed in Table 1 along with the scatter around the best fitting line. The uncertainties of the best fit param-

eters, also reported in the table, have been obtained by imposing $\chi^2_{\rm d.o.f.} = 1$.

As can be seen in Figure 2, the best fits to the model agree well with that to the data, within the scatter. We note that, in all relations plotted, the scatter in the model is larger than that of the real data and also larger than the internal scatter observed in similar relations obtained from the recent hydrodynamical simulations of galaxy mergers (see e.g. Hopkins et al. 2007a). However, we notice that a large fraction of our model BHs are found in low-mass systems for which the scatter in the scaling relation is large. On the contrary, in the real datasets (and hydro-simulations) the majority of BHs belong to massive galaxies for which, according to our model, the scatter in the scaling relation is significantly smaller. To investigate whether the difference in the intrinsic scatter is real or is induced by a different sampling of the BH population, for each BH scaling relation we have discretized the range of the observed host galaxy properties in finite bins and generated 500 sub-samples by randomly extracting $N_{\rm obs}(\Delta_X)$ model BHs from the parent catalogue, where $N_{\rm obs}(\Delta_X)$ is the number of BHs in the real dataset in each bin Δ_X. We have repeated the same fitting procedure in the 500 sub-samples and found that the scatter is significantly reduced in this exercise, as indicated in the last column of Table 1, that lists the average scatter in the sub-catalogues. Therefore, the mismatch in the scatter results from sampling different BH populations: small objects in the model, massive objects in the observations. Moreover, for the $\rm M_{BH}-\sigma_c$ relation the scatter is very close to 0.21, which is the value measured by Hopkins et al. (2007a) both in the observed and simulated data.

3.1.2 *Non-linear fits*

The agreement between model and data is satisfactory. However, we need to keep in mind that the model predictions for $V_{\rm c}$ and $\sigma_{\rm c}$ and the observed relation between $\log(M_{\rm BH})$ and $\log(M_{\rm DM})$ have been obtained assuming further theoretical hypotheses. Consequently, the more constraining and reliable relations are the ones between the BH masses and the bulge magnitudes and masses. Focusing on these relations and thanks to the huge number of model BHs, we have been able to investigate whether a non-linear fit provides a better match to the data. We find that the best fit is a quadratic function, $y = \alpha + \beta \cdot x + \gamma \cdot x^2$. Figure 3 shows this fit (heavy green lines), together with the medians, the first and third quartiles (black points with error bars) of the model output, computed in a discrete number of bins. The internal scatter is significantly smaller than in the linear fit case. The values of the best fit parameters are reported in Table 2. While we predict, on average, too low BH masses for a fixed $M_{\rm B}$ with respect to the observations (still consistent within the errors) the model predictions are in very good agreement with the data for the $\log(M_{\rm BH})-M_{\rm K}$ and $\log(M_{\rm BH})-\log(M_{\rm bulge})$ relations. Interestingly, the 3-parameters fit of the latter relation is in excellent agreement with the one found by Wyithe (2006) (magenta solid line in lower panel of Figure 3).

3.1.3 *The fundamental plane relation*

In Figure 4 we compare the BH fundamental plane relation of our model at different redshifts with that obtained by Hopkins et al. (2007a) using both observational data and the outputs of hydrodynamical simulations of galaxy mergers:

$$\log(M_{\rm BH}/M_\odot) = 7.93 + 0.72\log(M^*_{11}) + 1.4\log(\sigma_{200}),$$

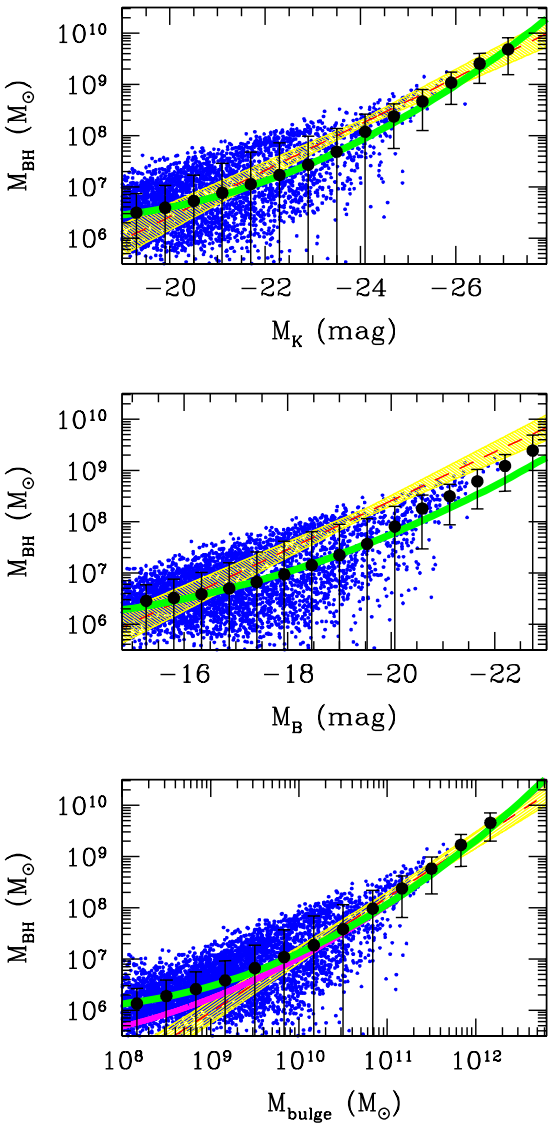

Figure 3. The tree model scaling relations best constrained by observations. Here the black dots (with error bars) represent the medians (and the corresponding first and third quartiles) of the model outputs, computed in a discrete number of bins. The green lines show the best three-parameters fits to the model outputs (blue points). The magenta line in the lower panel refers to the best-fit relation obtained by Wyithe (2006).

where M^*_{11} is the galaxy stellar mass in units of $10^{11}M_\odot$, and σ_{200} is the bulge velocity dispersion in units of 200 km s^{-1}. The red lines, bisectors of the plots, show the fundamental plane relation proposed by Hopkins et al. (2007a). Model prediction are represented by blue dots, the black line is the best fit to the model and the shaded area its 1σ scatter. At low redshifts the agreement is very good. This is not surprising since at $z \sim 0$ our model agrees with the $\rm M_{BH}-M_{bulge}$ and $\rm M_{BH}-\sigma_c$ scaling relations that represent fundamental plane projections. A discrepancy appears at high redshifts. However, at $z > 3$ the fit involves only few objects and therefore may not be very significant, especially when we account for the non-zero intrinsic scatter in the fundamental plane proposed by

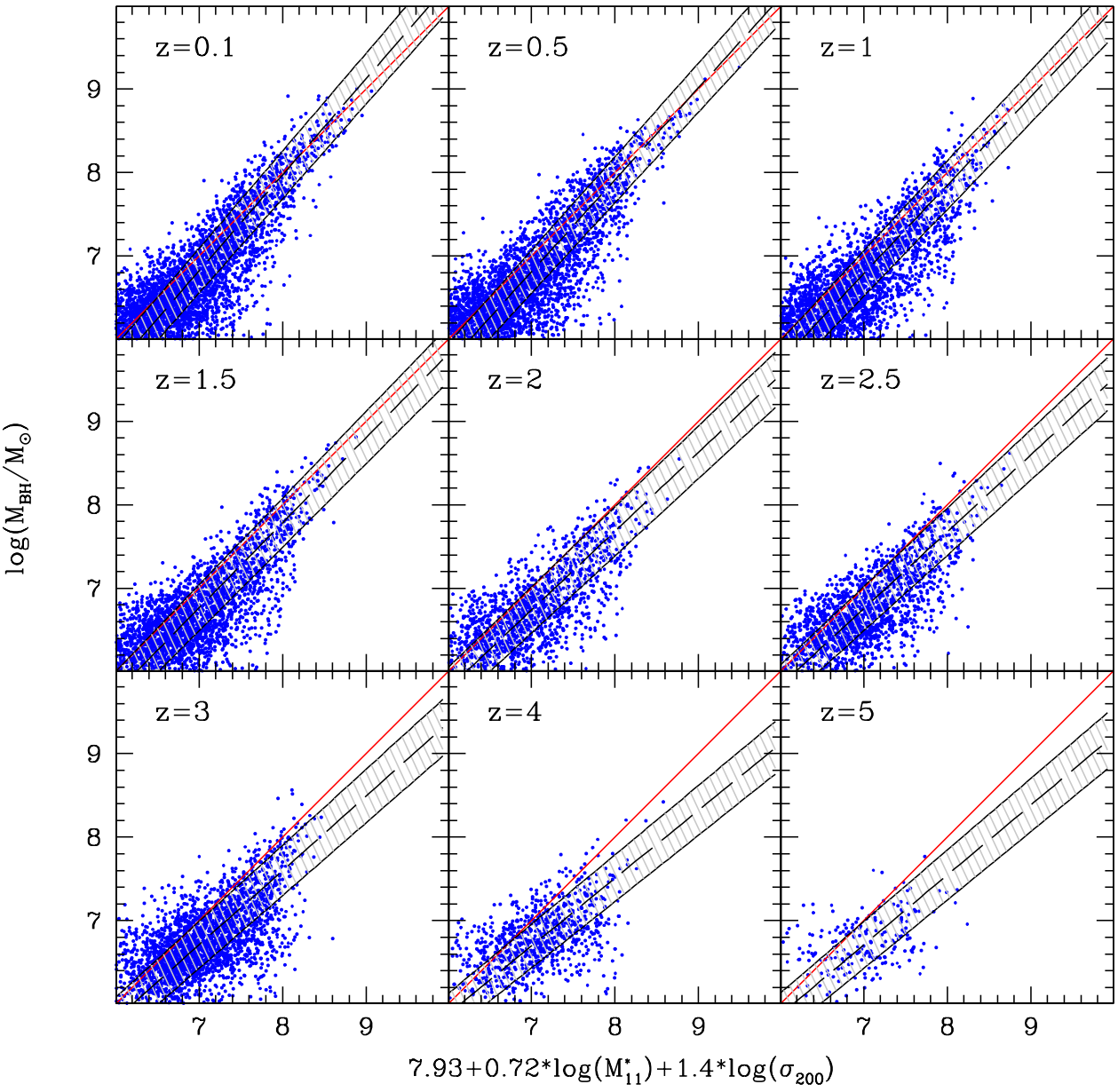

Figure 4. The BH fundamental plane in the redshift range $0.1 \leqslant z \leqslant 5$. The blue dots are the model outputs, while the grey shaded areas show the best-fits to them. The red lines, corresponding to the bisectors of the plots, are the predictions of Hopkins et al. (2007a). The galaxy stellar mass, M^*_{11}, is given in units of $10^{11} M_\odot$, while the bulge velocity dispersion, σ_{200}, is in units of 200 km s^{-1}.

Hopkins et al. (2007a). A remarkable success of our model is that it predicts very little evolution of the fundamental plane relation, at least out to $z = 3$, in agreement with Hopkins et al. (2007a). The intrinsic scatter, which does not evolve with time either, is 3 times larger than in Hopkins et al. (2007a) (we found a value around 0.6 at all redshifts, while the one reported by Hopkins et al. (2007a) is about 0.2). As discussed previously, the mismatch is reduced when using a number of model BHs consistent with the observed one.

3.1.4 Dependence on the accretion history

All scaling relations predicted by our model assume that BHs accrete mass instantaneously after merging events. What happens if we relax this assumption and specify the mass accretion rate instead? Figure 5 shows the impact of adopting different accretion recipes on the $M_{BH} - M_{bulge}$ relation. As usual, filled dots represent model predictions, grey shaded areas show the linear fit to the *DeLucia2006a* model scaling relation and the other hatched areas indicate the linear fit to the model predictions obtained with our

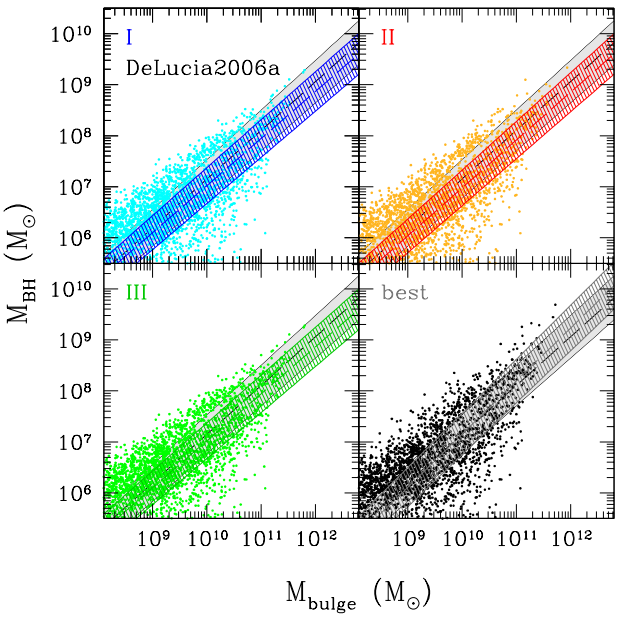

Figure 5. The $\log(M_{\rm BH}) - \log(M_{\rm bulge})$ scaling relation for our different prescriptions for the BH mass accretion. The filled dots represent model predictions, the grey shaded areas show the linear fit to the *DeLucia2006a* model scaling relation and the other hatched areas indicate the linear fit to the *I*, *II* and *III* lightcurve models, as indicated by the labels. The black dots and grey shaded areas, in the lower right panel, show the prediction obtained with the parameterization given by the equations (21), as explained in Section 3.3.

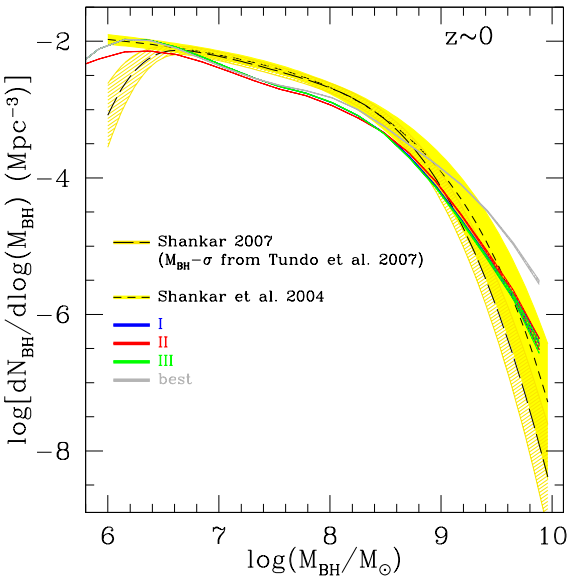

Figure 6. Comparison of the BH mass function predicted by lightcurve models *I*, *II* and *III* with the one observationally derived by Shankar et al. (2004), and with the new one obtained by Shankar (2007, in preparation) using the $M_{\rm BH} - \sigma$ relation by Tundo et al. (2007). The grey areas show the prediction obtained with the parameterization given by the equations (21), as explained in Section 3.3.

different recipes, as indicated by the labels[1]. Clearly, these predictions depend little on the assumed mass accretion histories for each individual quasar event (the fit parameters have fluctuations of no more than about 1%). This is a consequence of the fact that the BH scaling relations depend mainly on the total mass accreted, and very little on the time spent in the accretion process. We have verified that all other scaling relations, including also the fundamental plane relation, does not change significantly by adopting any of the mass accretion prescriptions described in Section 2.3.3.

3.2 The BH mass function

The BH mass function (MF) is defined as the differential comoving number density of BHs as a function of their mass. In Figure 6, we compare the BH MF predicted by our model for the prescriptions *I* (blue line), *II* (red) and *III* (green) with those observed by Shankar et al. (2004) (grey area) and by Shankar (2007, in preparation) (yellow area) at $z \sim 0$. In neither case the BH masses were determined directly: Shankar et al. (2004) derive the BH mass from the observed $M_{\rm BH} - L_{\rm bulge}$ relation while Shankar (2007) use the $M_{\rm BH} - \sigma_c$ relation of Tundo et al. (2007). We note that the model BH MF is in good agreement with the observed ones, within the mass range accessible to observations exept in the interval $\sim 10^7 - 10^9 M_\odot$, in which the number density of model BHs is smaller than the observed one.

The reason of the small mismatch between the observed and the model BH MFs will be investigated in a forthcoming paper in which we study the redshift evolution of the BH MF and its dependency on the properties of the host galaxy. Finally, we note that, as shown in Figure 6, the model predictions for the BH MF are robust with respect to the prescription adopted for the mass accretion history of the individual quasar episodes.

3.3 The AGN bolometric luminosity function

The luminosity function (LF) of AGN, namely the derivative of their co-moving number density with respect to luminosity, represents a unique tool to understand their cosmological evolution. Semi-analytic models predict the total (bolometric) luminosity of a statistically complete AGN catalogue, and to compare model LFs with observations we need to specify a bolometric correction, i.e. how to convert the luminosities observed in a particular band into bolometric ones (Elvis et al. 1994; Marconi et al. 2004; Hopkins et al. 2006). Another correction is required to account for possible incompleteness effects (see e.g. Comastri 2004; Gilli et al. 2007), which includes the possible existence of a population of obscured AGN whose fraction may depend on the wavelength band and redshift (Elvis et al. 1994; Marconi et al. 2004; La Franca et al. 2005; Lamastra et al. 2006).

Here we compare our predictions with the bolometric LF obtained by Hopkins et al. (2007) from the LFs observed in different bands: radio (see e.g. Nagar et al. 2005), optical (see e.g. Kennefick et al. 1995; Schmidt et al. 1995; Koehler et al. 1997; Grazian et al. 2000; Fan et al. 2001; Wolf et al. 2003; Hunt et al. 2004; Cristiani et al. 2004; Croom et al. 2005; Richards et al. 2005, 2006; Siana et al. 2006; Fontanot et al. 2007; Shankar & Mathur 2007; Bongiorno et al. 2007), infra-red (see e.g. Brown et al. 2006; Matute et al. 2006; Babbedge et al. 2006), soft X-ray (see e.g. Miyaji et al. 2000, 2001; Silverman et al.

[1] The meaning of the black dots and shaded areas in the bottom left panel of Fig. 5 is discussed in Section 3.3.

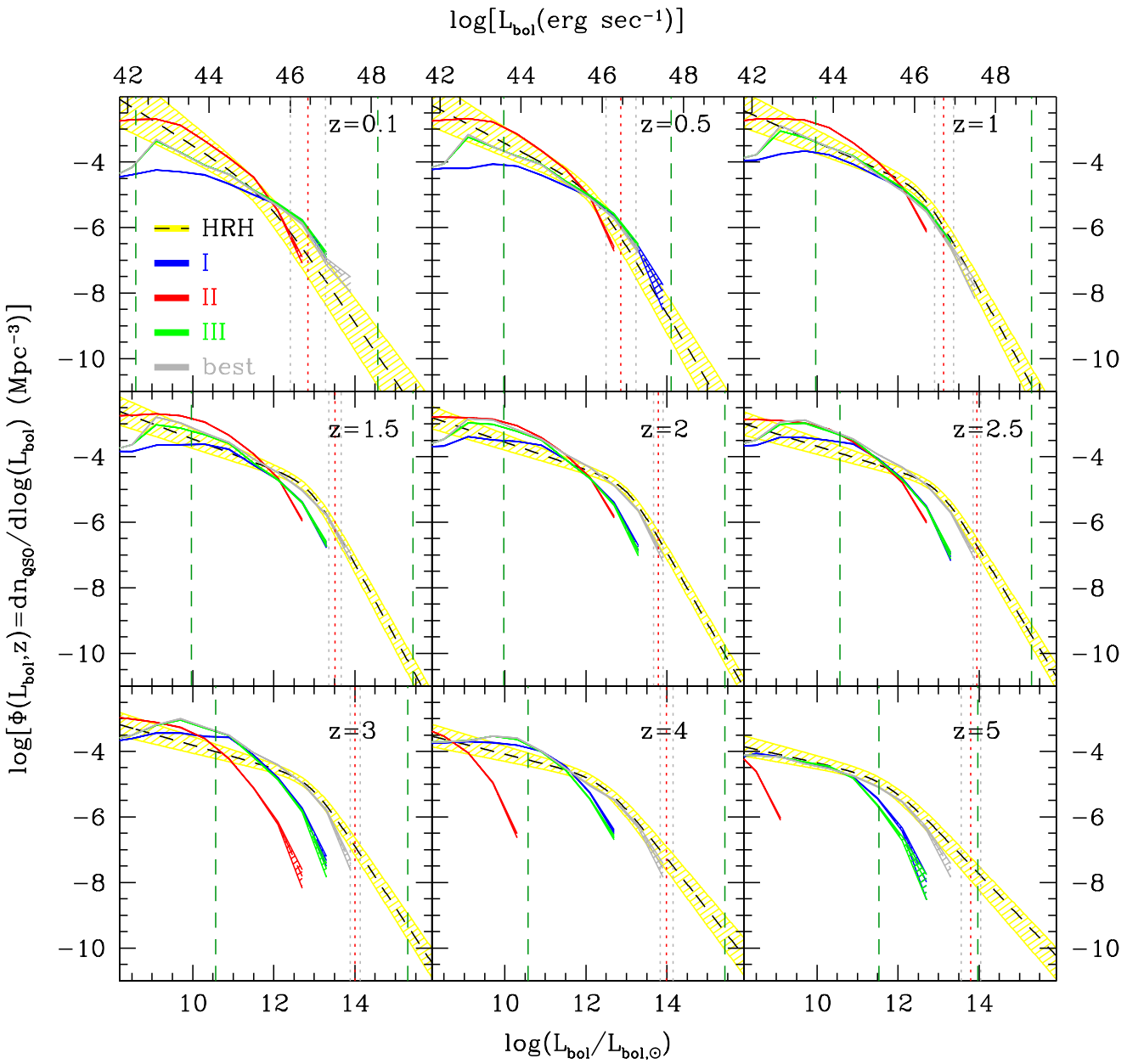

Figure 7. The bolometric LFs predicted by our lightcurve models *I* (blue bands), *II* (red bands) and *III* (green bands), in the redshift range $0.1 \leqslant z \leqslant 5$, are here compared with the best-fits to observational data obtained by Hopkins et al. (2007) (yellow bands). The grey areas show the predictions obtained with the parameterization given by the equations (21), as explained in Section 3.3. Uncertainties in the model LFs are computed by assuming Poisson statistics. The dashed vertical green lines mark the range of the bolometric luminosities accessible to observations. The dotted red vertical lines show the luminosities beyond which the LF of Hopkins et al. (2007) predicts a number of AGN in the whole volume of our simulation smaller than 10. The vertical grey dotted lines around the red ones have been calculated considering the error in the best-fit of Hopkins et al. (2007).

2005b; Hasinger et al. 2005), hard X-ray (see e.g. Barger et al. 2003; Ueda et al. 2003; Barger et al. 2003; Nandra et al. 2005; Sazonov & Revnivtsev 2004; Silverman et al. 2005a; La Franca et al. 2005; Shinozaki et al. 2006; Beckmann et al. 2006), and from emission lines (see e.g. Hao et al. 2005). Uncertainties in these corrections contribute to the scatter in the observed LF, i.e. to the width of the yellow areas in Figure 7 that show the AGN bolometric LF of Hopkins et al. (2007) at different redshifts. The model predictions are also represented by areas with different colours, with a width corresponding to 1σ Poisson error bars. The vertical, green dashed lines bracket the bolometric luminosity range accessible to observations. The vertical, red dotted lines show the luminosities beyond which the LF of Hopkins et al. (2007) predicts less than 10 AGN in the volume of our simulation, i.e. the maximum luminosities at which our model BH sample

is statistically meaningful; 1σ uncertainties on this maximum luminosity are represented by the two grey dotted lines.

From Figure 7 we see that, on average, type-*I* lightcurve underestimates the AGN number density at all epochs. However, while at high redshifts the model matches the faint-end of the LF and underpredicts the number density of the bright objects, the situation is completely reversed at $z \sim 0$, where the model correctly reproduces the number density of bright AGN but underestimates the faint ones. At low redshifts the problem can be alleviated by reducing the Eddington factor, as in our type-*II* lightcurve. However, in this case the discrepancy between model and data at high redshifts increases. Adopting the type-*III* lightcurve allows to match observations in the whole range of luminosities in the redshift range $0.5 \lesssim z \lesssim 1$, but overestimates the number of luminous AGN at $z \lesssim 0.5$ and underestimates them at $z \gtrsim 1$.

Therefore, we conclude that in our present semi-analytical framework we can reproduce the observed AGN LF at low and intermediate redshifts. However, at $z \gtrsim 1$, we under-predict the number density of bright AGN, regardless of the BH mass accretion rate and light curve model assumed for each quasar episode. To investigate if it is possible to modify our prescription for the mass accretion to fit the AGN LF at all redshifts, we tried different values of $f_{\rm Edd}$ and ε as a function of t and $M_{\rm BH}$, within physically motivated ranges. Despite of the considerable freedom in choosing $f_{\rm Edd}(t, M_{\rm BH})$ we failed to find a model able to match simultaneously the observed BH scaling relations, the BH MF, and the AGN LF, especially at high redshifts. We also used different plausible values for the BH seed mass, and we still were not able to fit the high-z LF. We interpret this failure as an indication that our theoretical framework itself is inadequate to account fully successfully for the AGN phenomenon.

One possible way out is to modify the model assumptions for the efficiency of BH growth in the *quasar mode* following mergers at high z. A significant improvement of our results at high redshifts can for example be obtained by substituting equation (12) and (13) with

$$\begin{cases} f_{\rm BH} = 0.01 \cdot \log\left(\frac{M_{\rm BH}}{10^3 M_\odot} + 1\right) \cdot z & z > 1.5 \,{\rm and}\, M_{\rm BH} > 10^6 M_\odot \\ \Delta M_{\rm BH,Q} = 0.01 \cdot m_{\rm cold} & z > 6 \end{cases} \quad (21)$$

while keeping prescription *III* for the quasar lightcurves. The predictions of this new model for the $\log(M_{\rm BH}) - \log(M_{\rm bulge})$ scaling relation is shown as black dots in the bottom-right plot of Fig. 5. Model predictions for the BH MF and AGN LF are shown in Figures 6 and 7, respectively. An accretion efficiency that increases with the redshift has been already advocated in the *dynamical model* of Croton (2006). A physical justification to this assumption is provided by Mo et al. (1998). Indeed, their model predicts that galactic disks were more centrally concentrated in the past, making it more efficient the BH feeding at high redshift. It is worth stressing that equation (21) might not provide the best fit to the data as we did not explore the parameter space systematically. However, it suggests that a good match to the observed scaling relations, BH MF and AGN LFs can be obtained within our semi-analytic framework by modest changes of the BH growth at high redshifts. The solution provided by equation (21) is not unique either, since larger amounts of mass can be accreted also by invoking alternative mechanisms that trigger gas accretion episodes, for example by secular evolution through disk instabilities, or by alluding to a higher gas cooling efficiency (see e.g. Viola et al. 2007).

4 CONCLUSIONS

In this paper we have used and extended a semi-analytic model for the co-evolution of galaxies and their central BHs, developed on the outputs of the Millennium Simulation (Springel et al. 2005), and described in detail in Croton et al. (2006) and De Lucia & Blaizot (2007). The aim of the model is to reproduce the observed properties of BHs, AGN and their galaxy hosts. The physical assumptions in the model with respect to BH growth can be divided into two sets. The first one concerns the mass accretion history of the central BHs in halos, where we distinguish between *radio mode* and *quasar mode* (Croton et al. 2006). This set makes predictions for the relation between BH and galaxy host properties, which can be compared to the observed scaling relations between BH mass and different properties of their host galaxies. The second set of prescriptions specifies the detailed AGN activity and lightcurve of individual quasar episodes, and leads to predictions for the AGN LF as a function of redshift. We considered three different models for this detailed AGN activity, one of them motivated by the results of recent hydrodynamical simulations of galaxy mergers that include BH growth and feedback (Hopkins et al. 2005; Di Matteo et al. 2005; Springel et al. 2005).

The main results of our analysis are as follows:

(i) The semi-analytic model is approximately able to reproduce the observed BH scaling relations over the whole range of BH masses and galaxy properties probed by observations. The intrinsic scatter in the model is significantly larger than in the data, a mismatch that can in part be accounted for by adopting the observational selection criteria to obtain a mock BH catalogue with similar characteristics as the observed one.

(ii) We find evidence that a quadratic relationship provides a significantly better fit to some of the model scaling relationships than a linear one, as already noticed by Wyithe (2006).

(iii) Our model also matches the BH fundamental plane relation derived by Hopkins et al. (2007a), and successfully predicts very little evolution of this plane, at least out to $z \sim 3$.

(iv) The model BH mass function is in good agreement with the observed one within the mass range accessible by observations, except on the range $\sim 10^7 - 10^9 M_\odot$, in which the number density predicted by the model is smaller than the observed one.

(v) Model predictions for the BH mass function, scaling relations and fundamental plane relation are basically unaffected when using different prescriptions for the AGN lightcurves of individual quasar events. This is because these predictions are only sensitive to the model assumptions for the absolute growth of the BHs in each merger event.

(vi) The AGN LF is systematically underestimated by assuming that BHs accrete mass with a constant Eddington factor $f_{\rm Edd} = 1$. The detail of the discrepancy, however, change with redshift since at high z the model matches the faint-end of the LF but underpredicts the number density of the brightest objects, while the situation is reversed at $z \sim 0$, in agreement with the results of several semi-analytic models (see, e.g. Marulli et al. 2007, and references therein). Reducing the Eddington ratio, as in our lightcurve model *II*, alleviates the faint-end mismatch but amplifies the bright-end discrepancy at high redshifts. A significant improvement at low redshifts is obtained when the Eddington-limited growth of the BH is followed by a long quiescent phase with lower Eddington ratios, as suggested by Hopkins et al. (2005) and implemented in our lightcurve model *III*. In this case our model is able to match the observed AGN LF in the interval $0.1 \lesssim z \lesssim 1$, over the whole range of luminosities that are accessible to observations and where our pre-

dictions are statistically significant. However, our predicted number density of bright AGN is still biased low at $z \gtrsim 1$.

(vii) Our model is able to account for all observations considered in this work apart for the AGN LF at high redshifts. We were not able to eliminate this mismatch by simply modifying the accretion efficiency, ϵ, the Eddington factor, $f_{\rm Edd}$, or the BH seed mass (when considered in physically plausible ranges). Clearly, we need to modify assumptions in the underlying semi-analytic framework for BH growth. A simple, *ad hoc* increase of the mass fraction accreted during the *quasar mode* at high redshift can indeed remedy the problem. However, this solution is not unique as several high-redshift modifications to the original model, like new mechanisms that trigger BH activity in addition to galaxy merging or more efficient gas cooling resulting in a larger reservoir of cold gas, can be advocated to bring the predictions in line with observations. However, it remains to be seen whether any of these alternatives is physically plausible.

(viii) Our model predictions at $z < 3$ are robust to changes in the assumed BH seed mass, but are sensitive to it at larger redshift. We will further explore this issue in a subsequent paper where we plan to study to what extent current observations can constrain the seed BH mass function.

From our analysis we conclude that the AGN LF at high redshifts constitutes a strong constraint for semi-analytic models that describe the co-evolution of galaxies, BHs and AGN. This suggests that significant improvements can be obtained in two ways. From the theoretical side, we need to develop a physically motivated mechanism capable of increasing the number density of bright AGN at $z \gtrsim 1$ without modifying the model predictions at low redshifts. From the observational point of view, we need to improve the AGN LF estimates at high redshift, both by enlarging current high-z AGN samples and by reducing the current uncertainties originating from bolometric and incompleteness corrections, in particular for the population of Compton Thick AGN. In addition, other observational tests should be performed, like the ability of our model to match the observed AGN clustering, as quantified by the angular and spatial two-point correlations function. In particular, Lidz et al. (2006) pointed out that the luminosity dependence of quasar clustering can discriminate between different lightcurve models, a question we will address in a forthcoming work.

ACKNOWLEDGMENTS

We thank Simon D. M. White, Gabriella De Lucia, Andrea Merloni, Philip Hopkins, Francesco Shankar and Umberto Maio for very useful discussions. FM thanks the Max-Planck-Institut für Astrophysik for the kind hospitality. We acknowledge financial contribution from contracts ASI-INAF I/023/05/0, ASI-INAF I/088/06/0 and INFN PD51. SB acknowledges the PhD fellowship of the International Max Planck Research School in Astrophysics, and the support received from a Marie Curie Host Fellowship for Early Stage Research Training

REFERENCES

Akritas M. G., Bershady M. A., 1996, ApJ, 470, 706
Babbedge T. S. R., et al., 2006, MNRAS, 370, 1159
Baes M., Buyle P., Hau G. K. T., Dejonghe H., 2003, MNRAS, 341, L44
Barger A. J., Cowie L. L., Capak P., Alexander D. M., Bauer F. E., Brandt W. N., Garmire G. P., Hornschemeier A. E., 2003, ApJ, 584, L61
Barger A. J., et al., 2003, AJ, 126, 632
Beckmann V., Soldi S., Shrader C. R., Gehrels N., Produit N., 2006, ApJ, 652, 126
Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38
Binney J., Tremaine S., 1987, Galactic dynamics. Princeton, NJ, Princeton University Press, 1987, 747 p.
Bongiorno A., et al., 2007, preprint, astro-ph/0704.1660, 704
Brown M. J. I., et al., 2006, ApJ, 638, 88
Cattaneo A., Blaizot J., Devriendt J., Guiderdoni B., 2005, MNRAS, 364, 407
Cavaliere A., Vittorini V., 2002, ApJ, 570, 114
Cimatti A., Daddi E., Renzini A., 2006, A&A, 453, L29
Ciotti L., Ostriker J. P., 2007, ApJ, 665, 1038
Colless M., et al., 2001, MNRAS, 328, 1039
Comastri A., 2004, in Barger A. J., ed., Astrophysics and Space Science Library Vol. 308 of Astrophysics and Space Science Library, Compton-Thick AGN: The Dark Side of the X-Ray Background. pp 245–+
Cowie L. L., Binney J., 1977, ApJ, 215, 723
Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839
Cox T. J., 2004, Ph.D. Thesis, University of California, Santa Cruz
Cristiani S., et al., 2004, ApJ, 600, L119
Croom S. M., et al., 2005, MNRAS, 356, 415
Croton D. J., 2006, MNRAS, 369, 1808
Croton D. J., et al., 2006, MNRAS, 365, 11
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101
De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, MNRAS, 366, 499
Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2007, ArXiv e-prints, 705
Di Matteo T., Springel V., Hernquist L., 2005, Nat, 433, 604
Efstathiou G., 1992, MNRAS, 256, 43P
Efstathiou G., Rees M. J., 1988, MNRAS, 230, 5P
Elvis M., et al., 1994, ApJS, 95, 1
Enoki M., Nagashima M., Gouda N., 2003, PASJ, 55, 133
Fabian A. C., Nulsen P. E. J., 1977, MNRAS, 180, 479
Fabian A. C., Sanders J. S., Allen S. W., Crawford C. S., Iwasawa K., Johnstone R. M., Schmidt R. W., Taylor G. B., 2003, MNRAS, 344, L43
Fan X., et al., 2001, AJ, 121, 54
Feoli A., Mele D., 2007, preprint, astro-ph/0703675
Ferrarese L., 2002, ApJ, 578, 90
Ferrarese L., Ford H., 2005, Space Science Reviews, 116, 523
Ferrarese L., Merritt D., 2000, ApJ, 539, L9
Fontanot F., Cristiani S., Monaco P., Nonino M., Vanzella E., Brandt W. N., Grazian A., Mao J., 2007, A&A, 461, 39
Fontanot F., Monaco P., Cristiani S., Tozzi P., 2006, MNRAS, 373, 1173
Gebhardt K., et al., 2000, ApJ, 539, L13
Gilli R., Comastri A., Hasinger G., 2007, A&A, 463, 79
Gnedin N. Y., 2000, ApJ, 542, 535
Graham A. W., Driver S. P., 2007, ApJ, 655, 77
Graham A. W., Erwin P., Caon N., Trujillo I., 2001, ApJ, 563, L11
Granato G. L., et al., 2004, ApJ, 600, 580

Grazian A., Cristiani S., D'Odorico V., Omizzolo A., Pizzella A., 2000, AJ, 119, 2540
Haehnelt M. G., Rees M. J., 1993, MNRAS, 263, 168
Haiman Z., Loeb A., 1998, ApJ, 503, 505
Haiman Z., Menou K., 2000, ApJ, 531, 42
Hao L., et al., 2005, AJ, 129, 1795
Häring N., Rix H.-W., 2004, ApJ, 604, L89
Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417
Hatziminaoglou E., Mathez G., Solanes J.-M., Manrique A., Salvador-Solé E., 2003, MNRAS, 343, 692
Hoeft M., Yepes G., Gottlöber S., Springel V., 2006, MNRAS, 371, 401
Hopkins P. F., Hernquist L., Cox T. J., Keres D., 2007b, ArXiv e-prints, 706
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Di Matteo T., Springel V., 2006, ApJ, 639, 700
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007a, preprint, astro-ph/0701351
Hopkins P. F., Hernquist L., Martini P., Cox T. J., Robertson B., Di Matteo T., Springel V., 2005, ApJ, 625, L71
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
Hunt M. P., Steidel C. C., Adelberger K. L., Shapley A. E., 2004, ApJ, 605, 625
Kauffmann G., 1996, MNRAS, 281, 475
Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576
Kennefick J. D., Djorgovski S. G., de Carvalho R. R., 1995, AJ, 110, 2553
Kennicutt Jr. R. C., 1998, ApJ, 498, 541
Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2
Koehler T., Groote D., Reimers D., Wisotzki L., 1997, A&A, 325, 502
Kormendy J., Richstone D., 1995, ARA&A, 33, 581
Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482
La Franca F., et al., 2005, ApJ, 635, 864
Lamastra A., Perola G. C., Matt G., 2006, A&A, 449, 551
Lemson G., Virgo Consortium t., 2006, preprint, astro-ph/060801
Li Y., Hernquist L., Robertson B., Cox T. J., Hopkins P. F., Springel V., Gao L., Di Matteo T., Zentner A. R., Jenkins A., Yoshida N., 2007, ApJ, 665, 187
Lidz A., Hopkins P. F., Cox T. J., Hernquist L., Robertson B., 2006, ApJ, 641, 41
Lynden-Bell D., 1969, Nat, 223, 690
Magorrian J., et al., 1998, AJ, 115, 2285
Maio U., Dolag K., Ciardi B., Tornatore L., 2007, MNRAS, 379, 963
Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2006, preprint, astro-ph/0607424
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
Martini P., Weinberg D. H., 2001, ApJ, 547, 12
Marulli F., Branchini E., Moscardini L., Volonteri M., 2007, MNRAS, 375, 649
Marulli F., Crociani D., Volonteri M., Branchini E., Moscardini L., 2006, MNRAS, 368, 1269
Matute I., La Franca F., Pozzi F., Gruppioni C., Lari C., Zamorani G., 2006, A&A, 451, 443
McLure R. J., Dunlop J. S., 2002, MNRAS, 331, 795
McNamara B. R., Nulsen P. E. J., Wise M. W., Rafferty D. A., Carilli C., Sarazin C. L., Blanton E. L., 2005, Nat, 433, 45
Miyaji T., Hasinger G., Schmidt M., 2000, A&A, 353, 25
Miyaji T., Hasinger G., Schmidt M., 2001, A&A, 369, 49
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Morandi A., Ettori S., 2007, MNRAS, pp 774–+
Nagar N. M., Falcke H., Wilson A. S., 2005, A&A, 435, 521
Nandra K., Laird E. S., Steidel C. C., 2005, MNRAS, 360, L39
Netzer H., Trakhtenbrot B., 2007, ApJ, 654, 754
Novak G. S., Faber S. M., Dekel A., 2006, ApJ, 637, 96
Percival W., Miller L., 1999, MNRAS, 309, 823
Peterson J. R., Paerels F. B. S., Kaastra J. S., Arnaud M., Reiprich T. H., Fabian A. C., Mushotzky R. F., Jernigan J. G., Sakelliou I., 2001, A&A, 365, L104
Richards G. T., et al., 2005, MNRAS, 360, 839
Richards G. T., et al., 2006, AJ, 131, 2766
Richstone D., et al., 1998, Nat, 395, A14+
Salpeter E. E., 1964, ApJ, 140, 796
Sánchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 189
Sazonov S. Y., Revnivtsev M. G., 2004, A&A, 423, 469
Schmidt M., Schneider D. P., Gunn J. E., 1995, AJ, 110, 68
Seljak U., 2002, MNRAS, 334, 797
Shankar F., Mathur S., 2007, ApJ, 660, 1051
Shankar F., Salucci P., Granato G. L., De Zotti G., Danese L., 2004, MNRAS, 354, 1020
Shinozaki K., Miyaji T., Ishisaki Y., Ueda Y., Ogasaka Y., 2006, AJ, 131, 2843
Siana B., et al., 2006, preprint, astro-ph/0604373
Sijacki D., Springel V., di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Silverman J. D., et al., 2005a, ApJ, 618, 123
Silverman J. D., et al., 2005b, ApJ, 624, 630
Soltan A., 1982, MNRAS, 200, 115
Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504
Spergel D. N., et al., 2003, ApJS, 148, 175
Spergel D. N., et al., 2007, ApJS, 170, 377
Springel V., 2005, MNRAS, 364, 1105
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., et al., 2005, Nat, 435, 629
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V., Yoshida N., White S. D. M., 2001, New Astronomy, 6, 79
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Tamura T., Kaastra J. S., Peterson J. R., Paerels F. B. S., Mittaz J. P. D., Trudolyubov S. P., Stewart G., Fabian A. C., Mushotzky R. F., Lumb D. H., Ikebe Y., 2001, A&A, 365, L87
Tremaine S., et al., 2002, ApJ, 574, 740
Tundo E., Bernardi M., Hyde J. B., Sheth R. K., Pizzella A., 2007, ApJ, 663, 53
Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886
Viola M., Monaco P., Borgani S., Murante G., Tornatore L., 2007, ArXiv e-prints, 710
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
Volonteri M., Sikora M., Lasota J.-P., 2007, preprint, astro-ph/0706.3900, 706
Wang J., De Lucia G., Kitzbichler M. G., White S. D. M., 2007, preprint, astro-ph/0706.2551, 706
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, MNRAS, 372, 1161
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wolf C., Wisotzki L., Borch A., Dye S., Kleinheinrich M., Meisenheimer K., 2003, A&A, 408, 499
Wyithe J. S. B., 2006, MNRAS, 365, 1082

Wyithe J. S. B., Loeb A., 2003, ApJ, 595, 614