A method for determining AGN accretion phase in field galaxies

Miroslav Micic,1✴ Nemanja Martinović1✴ and Manodeep Sinha2✴

1Astronomical Observatory Belgrade, Volgina 7, 11000, Belgrade, Serbia
2Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, Victoria 3122, Australia

ABSTRACT
Recent observations of active galactic nucleus (AGN) activity in massive galaxies (log $M_*/M_\odot > 10.4$) show the following: (1) at $z < 1$, AGN-hosting galaxies do not show enhanced merger signatures compared with normal galaxies, (2) also at $z < 1$, most AGNs are hosted by quiescent galaxies and (3) at $z > 1$, the percentage of AGNs in star-forming galaxies increases and becomes comparable to the AGN percentage in quiescent galaxies at $z \sim 2$. How can major mergers explain AGN activity in massive quiescent galaxies that have no merger features and no star formation to indicate a recent galaxy merger? By matching merger events in a cosmological $N$-body simulation to the observed AGN incidence probability in the COSMOS survey, we show that major merger-triggered AGN activity is consistent with the observations. By distinguishing between ‘peak’ AGNs (recently merger-triggered and hosted by star-forming galaxies) and ‘faded’ AGNs (merger-triggered a long time ago and now residing in quiescent galaxies), we show that the AGN occupation fraction in star-forming and quiescent galaxies simply follows the evolution of the galaxy merger rate. Since the galaxy merger rate drops dramatically at $z < 1$, the only AGNs left to be observed are the ones triggered by old mergers that are now in the declining phase of their nuclear activity, hosted by quiescent galaxies. As we go towards higher redshifts, the galaxy merger rate increases and the percentages of ‘peak’ AGNs and ‘faded’ AGNs become comparable.

Key words: black hole physics – surveys – galaxies: active – galaxies: haloes – galaxies: nuclei – dark matter.

1 INTRODUCTION
Galaxies residing outside galaxy clusters are known as field galaxies. Their name implies a certain level of isolation, either in time between major interactions (∼3 Gyr: Verdes-Montenegro et al. 2005) or through the surrounding environmental density (Dressler 1980).

The topic of this article is active galactic nucleus (AGN) activity in field elliptical galaxies. These are massive (log $M_*/M_\odot > 10.4$) galaxies, thought to be formed in gas-rich major mergers of disc/spiral galaxies (Toomre 1977; White 1978, 1979; Gerhard 1981; Negroponte & White 1983; Barnes 1988; Hernquist 1989; Barnes & Hernquist 1996; Naab, Jesseit & Burkert 2006; Novak et al. 2012).

We focus on field AGNs, because AGNs in massive elliptical galaxies are field phenomena. Hwang et al. (2012) have studied a sample of almost a million Sloan Digital Sky Survey (SDSS) galaxies. They found a factor of three larger AGN fraction in the field compared with clusters. At higher redshift, this increase is even more pronounced. Martini, Sivakoff & Mulchaey (2009) found an increase by a factor of 8 at redshift $z = 1$.

Galaxy mergers are also field phenomena. Low velocity dispersion in galaxy groups in the field leads to ‘slow encounters’ (Binney & Tremaine 1987), which are necessary for the merger to occur. ‘Fast encounters’ are a characteristic of galaxy clusters. Energy input and dynamical friction scale as $v^{-2}$ (Binney & Tremaine 1987) and do not lead to a merger but instead small perturbations, which can fuel a low-luminosity AGN (Lake, Katz & Moore 1998).

For a long time, major galaxy mergers have been the main mechanism for driving AGN activity (Sanders et al. 1988; Barnes & Hernquist 1996; Cavaliere & Vittorini 2000; Menci et al. 2004, 2006; Croton et al. 2006; Hopkins et al. 2006; Raferty et al. 2011; Hopkins 2012). Observational evidence indicates post-merger features in galaxies hosting AGNs and quasars (Surace & Sanders 1999; Canalizo & Stockton 2000, 2001; Surace, Sanders & Evans 2000), lending credence to the theoretical picture of mergers as drivers of AGN activity. Fiore et al. (2012) found that theoretical models using galaxy interactions as the AGN triggering mechanism are able to reproduce high-redshift ($z = [3, 7]$) AGN
luminous AGNs hosting SMBHs with masses $\geq 10^7 M_{\odot}$. Otherwise, the AGN is observed in the declining phase of its nuclear activity. This would place it in a massive, red, elliptical galaxy long after merger features can be detected, but its activity would still be consistent with a merger-driven model.

AGN hosts in the COSMOS survey are mainly massive, red galaxies. Hence, their AGNs could potentially be merger-driven, past their peak activity during the ‘green valley’ and in the declining ‘red sequence’ phase. This interpretation would be consistent with the merger-driven scenario for AGN activity and with the recent Schawinski et al. (2014) scheme for galaxy evolution.

We describe our method in Section 2 and introduce two models based on the initial BH mass function. In Section 3, we present our best-fitting model. In Section 4, we determine the phase of AGN activity in the COSMOS survey. We discuss the implications of our results in Section 5.

2 METHOD

The three major components in our model are as follows: dark matter halo (DMH) merger trees from a cosmological N-body simulation, the Shen (2009) fit-by-observations semi-analytical model for major merger-driven growth of SMBHs and the Bongiorno et al. (2012) study of $\sim 1700$ AGNs and their host galaxies in the COSMOS field survey.

The main idea is to track ‘field DMHs’ undergoing major mergers in our N-body cosmological simulation, use the Shen (2009) SMBH growth model and match it to the observations. We then overlay our field with the COSMOS field, match simulated galaxies to the observed COSMOS galaxies and assign COSMOS AGNs to them. Next, we find whether the observed AGNs are at their peak activity or in the declining phase.

Here is the outline of our model presented in Fig. 1.

(i) $z_{\text{initial}}$ is the redshift of a DMH merger. Haloes touch and the smaller halo is inside the larger halo at all later times.

(ii) $M_{H,1}$ and $M_{H,2}$ are the masses of merging haloes at $z_{\text{initial}}$. We consider major mergers only, when the mass ratio of merging haloes is $\geq 0.3$.

(iii) If a merging halo did not have a major merger in its history, we assume that the halo hosts a spiral galaxy. If it had a major merger previously, we assume it hosts an elliptical galaxy.

(iv) We seed spiral galaxies with pristine BHs ($\sim 10^{-3} - 10^6 M_{\odot}$) and elliptical galaxies with BHs from the Ferrarese (2002) $M_{\text{BH}} - M_{\text{DMH}}$ relation. Masses of BHs hosted by merging haloes are $M_{\text{BH},1}$ and $M_{\text{BH},2}$. Mergers of two elliptical galaxies do not trigger AGN activity (dry mergers).

(v) $z_{\text{AGN}}$ is the redshift at which a smaller halo cannot be identified any more inside the larger one, which means that the merger of DMHs has finished. We assume that the mergers of galaxies and black holes have finished too and that accretion on to the new SMBH starts and enters an AGN phase, which has pre-peak and peak activity.

(vi) Even before the central BHs ($M_{\text{BH},1}$ and $M_{\text{BH},2}$) form a binary (BHB), they accrete at a rate of $\sim 10^{-3} - 10^{-4} M_{\odot} \text{yr}^{-1}$ (Capelo et al. 2015), from $z_{\text{initial}}$ at $R_{\text{separation}}$ until a BHB forms at $\lesssim \text{kpc}$ distance. This is the ‘pre-BHB accretion’ phase.

(vii) The SMBH mass entering the AGN phase at $z_{\text{AGN}}$ is then $M_{\text{BH,initial}} = M_{\text{BH},1} + M_{\text{BH},2}$.
M. Micic, N. Martinović and M. Sinha

Figure 1. Graphical sketch showing the main steps of the methodology.

(viii) $M_{\text{H, AGN}}$ is the mass of the DMH hosting an AGN at $z_{\text{AGN}}$. We adopt the Shen (2009) model for AGN activity in field galaxies. This model implies a halo mass range $3 \times 10^{11} h^{-1} M_\odot < M_{\text{H, AGN}} < 10^{12} (1+z_{\text{AGN}})^{3/2} h^{-1} M_\odot$.

(ix) $M_{\ast, \text{AGN}}$ is the mass of the galaxy hosting an AGN at $z_{\text{AGN}}$. We consider only AGNs in massive galaxies ($\log(M_{\ast, \text{AGN}}/M_\odot) > 10.4$). The galaxy mass is obtained by using Rodriguez-Puebla et al. (2015)'s $M_{\ast} - M_{\text{DMH}}$ relation.

(x) $P(L_X)$ is the probability of a galaxy hosting an AGN of a given luminosity at $z_{\text{AGN}}$ (Bongiorno et al. 2012). From it, we obtain $L_X$ and calculate the bolometric AGN luminosity $L_{\text{COSMOS}}$.

(xi) $M_{\text{BH, predicted}}$ is the SMBH mass predicted by the Shen (2009) model, where $M_{\text{BH, initial}}$ is the input parameter given in point (vii) and the peak luminosity is replaced by $L_{\text{COSMOS}}$.

(xii) $z_{\text{final}}$ is the redshift of the post-merger halo $M_{\text{H, final}}$.

(xiii) $M_{\text{BH, final}}$ is the 'true' mass of the post-merger BH, derived from the $M_{\text{BH}} - \sigma_{\text{sph}}$ relation, calibrated to the local $M_{\text{BH}} - M_{\text{DMH}}$ relation (Ferrarese 2002).

(xiv) If the observed COSMOS AGN is at peak activity, then $M_{\text{BH, predicted}}$ has to be at least as large as $M_{\text{BH, final}}$. Otherwise, the AGN is in the declining phase of its nuclear activity.

In the following sections we describe the simulation, data and modelling in more detail.

2.1 Cosmological N-body simulation

Using GADGET2 (Springel, Yoshida & White 2001; Springel et al. 2005), we performed a high-resolution cosmological N-body simulation within a comoving periodic box with size $130 \, \text{Mpc}^3$. WMAP5-like (Komatsu et al. 2009) cosmology was used ($\Omega_M = 0.25$, $\Omega_X = 0.75$, $n_s = 1$, $\sigma_8 = 0.8$ and $h = 0.7$) from $z = 599$ to $z = 0$ (84 snapshots). Initial conditions were computed with the 2LPT code (Crocce, Pueblas & Scoccimarro 2006). The simulation utilizes $512^3$ dark matter particles for a mass resolution of $1.14 \times 10^9 M_\odot$.

We generated halo catalogues using Robust Overdensity Calculation using K-space Topologically Adaptive Refinement (ROCKSTAR; Behroozi, Wechsler & Wu 2013b). ROCKSTAR combines friends-of-friends (FOF), phase-space and spherical overdensity analysis in locating haloes. Please see Behroozi et al. (2013a) for details on the ROCKSTAR algorithm. The merger tree was generated using CONSISTENT MERGER TREE (Behroozi et al. 2013a), a software package that is complementary with the ROCKSTAR halo finder.

2.2 AGNs and galaxies in the COSMOS survey

Bongiorno et al. (2012) have studied $\sim 1700$ AGNs in the COSMOS field obtained by combining X-ray and optical spectroscopic selections. This is a highly homogeneous and representative sample of obscured and unobscured AGNs over a wide redshift range ($0 < z < 4$). By using a spectral energy distribution (SED) fitting procedure, they have managed to separate host galaxy properties including the total stellar mass of galaxies hosting AGNs. One of their results is the probability of a galaxy hosting an AGN of a given luminosity ($P(L_X)$) as a function of stellar mass in three redshift bins: $[0.3–0.8]$, $[0.8–1.5]$ and $[1.5–2.5]$ (fig. 14 in their article, hereafter F14). They grouped AGNs into four logarithmic X-ray ($2–10 \, \text{keV}$) luminosity bins: $[42.8–43.5]$, $[43.5–44.0]$, $[44.0–44.5]$ and $[44.5–46.0]$ in erg s$^{-1}$ units. They showed that, for a fixed mass range, observed field galaxies are more likely to host less luminous AGNs. The probability that a field galaxy hosts an AGN decreases with increasing AGN luminosity.

2.3 SMBH growth model

We adopt the Shen (2009) model for the hierarchical growth and evolution of SMBHs, assuming that AGN activity is triggered in...
major mergers. This model uses a general form of light curve, where a BH first grows exponentially at a constant-luminosity Eddington ratio of $\lambda_0 = 3$ (Salpeter 1964) to $L_{\text{peak}}$ at $t = t_{\text{peak}}$ and the luminosity then decays monotonically as a power law (Yu & Lu 2008).

Shen (2009) uses a variety of observations. The model adopts Hopkins, Richards & Hernquist (2007) compiled AGN bolometric luminosity function and both the observed redshift evolution and luminosity served Eddington ratio distributions.

The model successfully reproduces the observed AGN luminosity function and both the observed redshift evolution and luminosity dependence of the linear bias of AGN clustering.

The input parameters for the Shen (2009) model are $M_{\text{BH, initial}}$ (mass of the BH entering AGN phase), $L_{\text{peak}}$ (peak bolometric AGN luminosity) and $M_{\text{BH, relic}}$ (BH mass in the post-merger halo $M_{\text{DMH, post}}$ immediately after the AGN phase). To match our nomenclature, we have renamed $M_{\text{BH, relic}}$ to $M_{\text{BH, final}}$.

In our model, values for the first parameter come from the numerical simulation combined with semi-analytical modelling (details in Section 2.4).

We calculate $M_{\text{BH, final}}$ (details in Section 2.6) from the $M_{\text{BH}}$–$\sigma_{\text{sph}}$ relation (Kormendy & Ho 2013), where $\sigma_{\text{sph}}$ is the velocity dispersion of the stellar spheroid. $\sigma_{\text{sph}}$ is correlated with $V_{\text{vir}}$ by a constant (Ferrarese 2002). We set this constant to a value that reproduces the $z = 0$ (Ferrarese 2002) $M_{\text{BH}}$–$M_{\text{DMH}}$ relation. Since the $M_{\text{BH}}$–$\sigma_{\text{sph}}$ relation is expected to be non-evolving, one can find $M_{\text{BH}}$ in $M_{\text{DMH, post}}$ at any redshift. The outcome of this procedure is that the BH mass in high-redshift haloes, right after the AGN phase, is overestimated. This is expected to occur as the BH grows first, followed by post-merger halo growth through minor mergers and diffuse matter accretion. As we go towards lower redshifts, DMH growth catches up to SMBH growth to reproduce the local Ferrarese relation. Hence, we consider $M_{\text{BH, final}}$ to be the ‘true’ final BH mass.

$L_{\text{peak}}$ is the peak bolometric luminosity (details in Section 2.7) in the Shen (2009) light-curve model, to set the value to that guarantees that the accretion on to $M_{\text{BH, initial}}$ produces $M_{\text{BH, final}}$. This is our $L_{\text{peak, true}}$ best-fitting model, which reproduces the observed AGN activity, luminosity function, duty cycle and bias factor.

After we demonstrate that our merger-driven model reproduces the observed AGN statistics, we test whether the observed AGN activity in the COSMOS survey corresponds to the peak or to declining activity. Now, instead of $L_{\text{peak, true}}$ values for the peak luminosity ($L_{\text{peak}}$) are retrieved from the probability of a galaxy hosting an AGN of a given luminosity ($P(L_X)$) in the COSMOS survey (details in Section 2.8). We use this probability to seed galaxies with AGNs in 40 000 Monte Carlo realizations and grow SMBHs according to the Shen (2009) model (details in Section 2.9). The result is the probability that SMBHs grown in COSMOS AGNs match the true SMBHs grown in our best-fitting model.

2.4 Haloes, galaxies, black holes: initial values

We start by identifying major merger events in the merger trees of our cosmological N-body simulation. We define the masses of merging haloes as $M_{\text{H,1}}$ and $M_{\text{H,2}}$ at the time of the merger $z_{\text{initial}}$ (Fig. 1). We also check whether merging haloes had major mergers previously. A DMH without previous major mergers is an ancient halo hosting a disc/spiral galaxy with a large cold gas reservoir and a central BH that most likely formed through direct collapse of a gas cloud (Bromm & Loeb 2003; Begelman, Volonteri & Rees 2006; Begelman, Rossi & Armitage 2008). The latest observations (Mortlock et al. 2011) have shown that BH seeds had to be massive ($\sim 10^{6} - 10^{7} M_\odot$) in order to grow $\sim 10^9 M_\odot$ BHs at redshift $z \sim 7$.

The initial mass function (IMF) and mass range of the seed BHs are unknown. These BHs settle at the centres of disc/spiral galaxies, but their masses do not correlate with any of the galaxy properties.

Growth of these initial BHs through accretion can occur even before they form a binary (BHB), during the early stages of the galaxy merger as galaxies go through subsequent pericentric passages (Capelo et al. 2015). As the major merger of galaxies proceeds, gravitational torques generate large-scale gas inflows that drive the gas down to a subpc scale, where it can be accreted by the BH. Hence, BHs grow by a modest amount even before they form a binary (pre-BHB accretion). Modelling of this growth is a subject for numerous numerical studies. However, limited resolution and disparate subgrid physics recipes led to very different estimates of BH accretion rates. The latest results (Hayward et al. 2014; Capelo et al. 2015) show that a BH accretes at the rate $10^{-5} - 10^{-3} M_\odot$ yr$^{-1}$ for $\sim 1$ Myr before the AGN phase.

After a BH binary forms, accretion increases as BHs sink to overcome the last couple of kpc between them. During these last $\sim 100$ Myr before the BH merger, BHs double their masses (Roskar et al. 2015; Tamburello et al. 2016). Assuming that the Salpeter time is $\sim 50$ Myr, the corresponding Eddington ratio is 0.35. This would mean that a $10^4 M_\odot$ BH accretes at a rate of 0.1 $M_\odot$ yr$^{-1}$, while a $10^5 M_\odot$ BH accretes at the rate of 1 $M_\odot$ yr$^{-1}$ (over 100 Myr). This ‘BHB accretion’ phase finishes with BH binary coalescence into a new BH, which enters an AGN phase.

The unknown IMF for BH seeds and the rate of ‘pre-BHB accretion’ are the main sources of uncertainty in our modelling. We overcome this issue by considering two models, with the idea of constraining the lower and upper ends of the possible initial BH mass. Our lower constraint model (M1) assumes a log-normal IMF for BH seeds in the interval $\log (M_{\text{BH}}/M_\odot) = [4.5, 5.5]$ centred at $10^5 M_\odot$ and pre-BHB accretion on to the BH at the centre of the DMH. Our higher constraint model M2, we set $\log (M_{\text{BH}}/M_\odot) = [5.0, 6.0]$ centred at $10^5 M_\odot$ and $\sim 10^4 M_\odot$ yr$^{-1}$. We seed DMHs with BHs by randomly choosing BH masses from these IMFs in Monte Carlo realizations. In the last 100 Myr before $z_{\text{AGN}}$, we double the BH mass (BHB accretion).

If a DMH already had a major merger in its history, we assume it hosts an elliptical galaxy. A BH at the centre of an elliptical galaxy scales with the properties of the stellar spheroid, but also with the mass of the host DMH. We use the Ferrarese (2002) $M_{\text{BH}}$–$M_{\text{DMH}}$ relation with $\pm 10$ per cent scatter to seed elliptical galaxies with central BHs. Mergers of two elliptical galaxies do not trigger AGN activity (dry mergers). If an elliptical galaxy merges with a spiral, there is no pre-BHB or BHB accretion on to the BH at the centre of the elliptical galaxy.

The initial mass of the BHs in both haloes ($M_{\text{BH,1}}$ and $M_{\text{BH,2}}$), combined with pre-BHB and BHB accretion (if the galaxy is spiral), produces $M_{\text{BH,1}}$ and $M_{\text{BH,2}}$. The initial BH mass that enters the AGN phase is then $M_{\text{BH, initial}} = M_{\text{BH,1}} + M_{\text{BH,2}}$.

2.5 AGN phase

At $z_{\text{AGN}}$, the initial BHs merge to form a new BH ($M_{\text{BH, initial}}$). The mass of the DMH hosting an AGN is then $M_{L_{\text{AGN}}}$. Accretion on to $M_{\text{BH, initial}}$ starts first with the pre-peak phase at the super-Eddington rate ($\lambda = 3$), followed by the declining phase, best described by fig. 2 in Shen (2009). We assume that the AGN reaches its peak activity at $z_{\text{AGN}}$. 


Note that in the Shen (2009) model, AGN activity starts at the time when haloes merge (not galaxies). Hence, the AGN activity in this model is pushed towards slightly higher redshifts. We find that the typical delay between a halo merger and the consequent galaxy merger is \( \Delta z = 0.2 \) in redshift space; this does not impact the overall results.

We adopt the Shen (2009) model for AGN activity in field galaxies. We consider haloes in the mass range \( 3 \times 10^{11} h^{-1} M_\odot < M_{\text{halo}} < 10^{12} (1 + z_{\text{agn}})^{1/3} h^{-1} M_\odot \). If the halo mass is too small, AGN activity cannot be triggered, while overly massive haloes cannot cool gas efficiently and BH growth halts (especially at low redshift; see Shen 2009). This excludes high-density environments (e.g. galaxy clusters) from our model and we are left with the AGN activity in the field. We do find that increasing the upper limit on host halo mass overpredicts the AGN luminosity functions at low redshift \( (z \lesssim 1) \).

The mass of the galaxy hosting an AGN is \( M_{\text{*,agn}} \). Since the topic of this article is to examine merger-driven AGN activity in massive galaxies in the field, we consider galaxies with \( \log(M_{\text{*,agn}}/M_\odot) > 10.4 \). In lower mass galaxies, SMBHs are more likely to accrete through secular processes related to channelling of the gas through bars or disc instabilities.

The galaxy mass is obtained by using Rodriguez-Puebla et al. (2015)’s \( M_{\text{DMH}} - M_{\text{halo}} \) relation for early-type (elliptical) galaxies (equations 17 and 18 and fig. 5 in their article) with scatter \( \sigma_M = \pm 0.136 \) dex (equation 37 in Rodriguez-Puebla et al. 2015). Scatter determines galaxy mass in every Monte Carlo realization.

### 2.6 Haloes, galaxies, black holes: final values

\( M_{\text{halo}} \) is the mass of the post-merger halo (immediately after the AGN phase) hosting the final (relic) SMBH. We chose to define the time \( z_{\text{agn}} \) to be \( \approx 100 \) Myr after AGN phase \( z_{\text{agn}} \), located in the first consecutive snapshot. However, the mass of the post-merger halo changes insignificantly in more than one snapshot after \( z_{\text{agn}} \). In fact, our results do not change even when we use \( M_{\text{halo}} \) instead of \( M_{\text{halo}} \). This occurs because, at the time of the galaxy merger, a new halo has already formed and for some time after the AGN phase it stays intact. Later, it continues growing by minor mergers and diffuse matter accretion. This implies that, at first, the mass of the final (relic) SMBH \( M_{\text{BH,final}} \) hosted by \( M_{\text{halo}} \) will be overestimated compared with the local Ferrarese \( M_{\text{BH,final}} - M_{\text{DMH}} \) relation. As \( M_{\text{halo}} \) grows in mass over time, \( M_{\text{BH,final}} - M_{\text{DMH}} \) final relation approaches the Ferrarese relation.

Since the \( M_{\text{BH}} - \sigma_{\text{g}} \) relation is expected to be non-evolving (Gaskell 2009; Shankar, Bernardi & Haiman 2009; Salvianter, Shields & Bonning 2015; Shen et al. 2015), one can find \( M_{\text{BH,final}} \) in \( M_{\text{DMH,final}} \) at any redshift. First, one can rewrite equation (3) in Ferrarese (2002) as

\[
\frac{V_{\text{vir}}}{200 \text{ km s}^{-1}} = \left( \frac{M_{\text{halo}}}{2.7 \times 10^{12} M_\odot} \right)^{1/3}.
\]

Next, \( \sigma_{\text{g}} = C \times V_{\text{vir}} \). From equation (7) in Kormendy & Ho (2013),

\[
M_{\text{BH,final}} = 0.309 \left( \frac{\sigma_{\text{g}}}{200 \text{ km s}^{-1}} \right)^{4.38},
\]

with scatter \( \sigma = \pm 0.28 \) dex.

We find that for \( C = 0.77 \), our \( M_{\text{BH,final}} - M_{\text{DMH,final}} \) relation at \( z = 0 \) matches the local Ferrarese relation. As we go towards higher redshifts, the Ferrarese relation evolves and \( M_{\text{BH,final}} \) is overestimated, while \( M_{\text{BH}} - \sigma_{\text{g}} \) does not evolve.

### 2.7 Finding the best-fitting model

Now that we have obtained \( M_{\text{BH,initial}} \) and \( M_{\text{BH,final}} \), we can calculate the \( L_{\text{peak}} \) necessary to produce \( M_{\text{BH,final}} \). As already mentioned in Section 2.5, we use the Shen (2009) AGN light curve with a pre-peak exponential growth phase followed by a post-peak power-law decline. To calculate \( L_{\text{peak}} \), we rewrite equation (29) in Shen (2009):

\[
L_{\text{peak}} = 3M_{\text{BH,final}}E_{\text{edd}} \left( 1 - \frac{2\ln f}{3} \right)^{-1}, \tag{3}
\]

and

\[
f = \frac{3E_{\text{edd}}M_{\text{BH,initial}}}{L_{\text{peak}}}, \tag{4}
\]

where \( E_{\text{edd}} = 1.26 \times 10^{53} \text{ erg s}^{-1} M_\odot^{-1}. \)

The descending phase is presented by equation (24) in Shen (2009):

\[
L(t) = L_{\text{peak}} \left( \frac{t}{t_{\text{peak}}} \right)^{-\alpha}, \tag{5}
\]

where \( \alpha = 2.5 \). Luminosities of all AGNs in all galaxies and at all redshifts decrease by three orders of magnitude from their peak luminosity in \( \sim 2 \) Gyr.

We use \( M_{\text{BH,initial}}, M_{\text{BH,final}} \) and \( L_{\text{peak}} \) to calculate the AGN luminosity function, active SMBH mass function, duty cycle and bias factor. We compare these with the observed values. We find that both M1 and M2 models reproduce observations without any additional modelling or parameter fixing. We continue with M1 and M2 as our best-fitting models and later replace \( L_{\text{peak}} \) with \( L_{\text{AGN}} \) to find the AGN activity phase in the COSMOS survey.

### 2.8 Matching COSMOS AGNs to \( M_{\text{*,agn}} \)

For \( \sim 1700 \) AGNs in the COSMOS field, Bongiorno et al. (2012) presented the probability of a galaxy of a certain mass hosting an AGN of given luminosity as a function of stellar mass in three redshift bins: \([0.3–0.8], [0.8–1.5]\) and \([1.5–2.5]\) (F14). They group AGNs into four X-ray (2–10 keV) luminosity bins in logarithmic space: \([42.8–43.5], [43.5–44.0], [44.0–44.5]\) and \([44.5–46.0]\) (erg s\(^{-1}\)) units.

The masses of their AGN-hosting galaxies are also separated into logarithmic bins: \([9.0, 10.0], [10.0, 10.4], [10.4, 10.7], [10.7, 10.9], [10.9, 11.2]\) (\( M_{\odot} \) units).

How to pick a luminosity from F14 and assign it to our \( M_{\text{BH,initial}} \)? We do this by grouping our simulated galaxies at the moment their \( M_{\text{BH,initial}} \) should start accreting.

We determine \( z_{\text{agn}} \), \( M_{\text{BH,initial}} \) and \( M_{\text{*,agn}} \) for every merger in our simulation and group them into redshift bins:

\[
\Delta z = [0.3–0.8; 0.8–1.5; 1.5–2.5]. \tag{6}
\]

and galaxy log-mass bins:

\[
\Delta M_{\ast} = [10.4–10.7; 10.7–10.9; 10.9–11.2]. \tag{7}
\]

since we study AGNs in massive galaxies only.

In this manner, we obtain nine \( \Delta z - \Delta M_{\ast} \) intervals. The number of galaxies belonging to each \( \Delta z - \Delta M_{\ast} \) interval is \( N_{\text{*,agn}} \). Next we match these \( \Delta z - \Delta M_{\ast} \) intervals to the \( \Delta z - \Delta M_{\ast} \) intervals in F14. According to F14, galaxies can host AGNs with luminosities in intervals

\[
\Delta L_X = [42.8–43.5; 43.5–44.0; 44.0–44.5; 44.5–46.0]. \tag{8}
\]

So the BHs in \( N_{\text{*,agn}} \) simulated galaxies can be assigned any of the luminosities from \( \Delta L_X \) intervals. How these luminosities should
be assigned is determined by the probability $P_{\text{AGN}, i}$ (data points in F14) defined for every $\Delta L_{X, i}$.

The $P_{\text{AGN}, i}$ in F14 tell us that every galaxy in a specific $\Delta z$-$\Delta M_*$ interval is more likely to host a low-luminosity AGN.

Since the number of galaxies in each $\Delta z$-$\Delta M_*$ interval is $N_{\text{AGN}}$, then the number of times a luminosity should be drawn from each luminosity interval $\Delta L_{X, i}$ is

$$N_{L, i} = \frac{P_{\text{AGN}, i}}{\sum P_{\text{AGN}, j}} \times N_{\text{AGN}}.$$  

(9)

The largest $N_{L, i}$ is for the interval $\Delta L_{X, i}$ with the smallest luminosities. The sum of $N_{L, i}$ is equal to $N_{\text{AGN}}$.

Next we randomly draw luminosities $L_{\text{AGN}, i}$ times from every corresponding $\Delta L_{X, i}$ and we randomly assign them to $N_{\text{AGN}}$ galaxies.

This is the first out of 40 000 Monte Carlo realizations where we draw luminosity values to be assigned to AGNs in each $\Delta z$-$\Delta M_*$ interval. Thus, for each $M_{\text{AGN}}$ we have a set of 40 000 COSMOS AGN luminosities. Since these are X-ray luminosities, we use equation (2) in Hopkins et al. (2007) to calculate bolometric luminosities. We address these luminosities as $L_{\text{COSMOS}}$. Next, we replace $L_{\text{peak}}$ in our best-fitting model with $L_{\text{COSMOS}}$.

### 2.9 Modelling SMBH growth in COSMOS

With the calculated $M_{\text{BH}, \text{initial}}$ (mass of the BH entering the AGN phase) and $L_{\text{COSMOS}}$ (COSMOS bolometric AGN luminosity), we have two input parameters for the Shen (2009) SMBH growth model. Evolution of AGN luminosities follows a universal general form of light curve, with initial exponential growth (pre-peak accretion) at constant Eddington ratio $\lambda = 3$ until it reaches $L_{\text{peak}}$, followed by a power-law decay. We replace $L_{\text{peak}}$ with $L_{\text{COSMOS}}$.

Note that there are two sets of 40 000 $M_{\text{BH}, \text{initial}}$ and $L_{\text{COSMOS}}$ for each $\Delta z$-$\Delta M_*$ interval, obtained from Monte Carlo realizations in two models: M1 (lower range of seed BH masses) and M2 (upper range of seed BH masses).

After applying the best-fitting parameters of Shen (2009) to their equation (29), the predicted SMBH mass ($M_{\text{BH, predicted}}$) after the AGN phase can be written as:

$$M_{\text{BH, predicted}} = \frac{L_{\text{COSMOS}}}{3L_{\text{Edd}}} \left( 1 - \frac{2\ln f}{3} \right)$$  

(10)

and

$$f = \frac{3L_{\text{Edd}}M_{\text{BH, initial}}}{L_{\text{COSMOS}}},$$  

(11)

where $L_{\text{Edd}} = 1.26 \times 10^{38}$ erg s$^{-1}$ $M_\odot^{-1}$.

Through Monte Carlo realizations, we take into account all possible seed values that could be assigned to the merging DMHs, all possible luminosities in each luminosity bin of Bongiorno et al. (2012), scatter in the Ferrarese (2002) $M_{\text{BH}}$-$M_{\text{DMH}}$ relation, scatter in the Shen (2009) $M_{\text{BH, final}}$-$M_{\text{DMH, post}}$ relation and scatter in the Rodriguez-Puebla et al. (2015) $M_{\text{DMH}}$-$M_{\text{DMH, residual}}$ relation. In the last mentioned scatter, the same halo can host a galaxy below or above $\log (M_*/M_\odot) = 10.4$. As the result, depending on the random draw from the scatter in each Monte Carlo realization, some haloes might drop from the analysis, while others might join.

At the end we have $M_{\text{BH, final}}$ from our best-fitting model and in 40 000 Monte Carlo realizations we produce the $M_{\text{BH, predicted}}$ in each $\Delta z$-$\Delta M_*$ interval. Now we can compare these two masses. If the observed AGN luminosities are indeed the peak luminosities at which most of the SMBH growth occurs, then the mass of the predicted SMBH should match the mass of the final SMBH. We calculate the percentage of realizations in which this condition is met.

### 3 RESULTS

#### 3.1 Best-fitting model

We apply a Shen (2009) major merger-driven AGN activity model to the merger trees in a cosmological N-body simulation. There are some differences between Shen (2009) and our model.

While Shen (2009) assumes a constant ratio of $10^{-3}$ between the initial and peak BH mass, we seed DMHs with BH seeds and follow their evolution before the AGN phase. Hence, the BH mass right before the AGN phase is not necessarily a constant fraction of the peak BH mass. Also, we calibrate the final SMBH mass at redshift $z = 0$ to the local Ferrarese relation. In this manner, the SMBH mass is overestimated at high redshift to accommodate late DMH evolution. In our model, super-Eddington accretion starts when galaxies merge, while in the Shen (2009) model the same occurs when DMHs merge.

Our best-fitting model for SMBH growth reproduces the observed AGN luminosity function, SMBH mass function, duty cycle and bias. Both M1 and M2 models can be considered as best-fitting models. The M2 model provides a slightly better fit to the observations, hence we show this match for the M2 model only.

Fig. 2 shows the BH mass function at three redshifts $z = [2.00, 1.25, 0.75]$. Horizontal and vertical bars show the full range for BH mass function in our Monte Carlo realizations, for active black holes only, in AGNs with log $L_X$ [erg s$^{-1}$] $\geq 43$, where $X = [2–10]$ keV. The dotted, blue line shows our BH mass function for all BHs. Overplotted as a thick black line is the active BH mass function for the same luminosity range from observations (HELLAS2XMM) of La Franca et al. (2005, presented in Fiore et al. 2012). Our best-fitting model follows the observed mass functions for active BHs. We slightly overestimate masses of active BHs at $z = 2$. This effect transfers to lower redshifts, where our BH mass function for all BHs (dotted, blue) slightly overpredicts the local BH mass function at $\sim 10^8$ $M_\odot$ (dashed, red, Merloni & Heinz 2008). At larger BH masses, our model underpredicts the local BH mass function. We find that this occurs due to the arbitrary cut-off at the higher mass end for haloes capable of hosting AGNs. When this upper limit for halo mass is doubled, we obtain a perfect match to the local BH mass function for $M_{\text{BH}} > 10^7$ $M_\odot$. However, at the same time, we overpredict the AGN luminosity function at $z < 1$ and log $L_X$ [erg s$^{-1}$] > 44 by a factor of 4. Similarly to Shen (2009), our model is incomplete at $M_{\text{BH}} \leq 10^5$ $M_\odot$ because we did not include contributions from AGNs triggered by secular processes or minor mergers.

Fig. 3 shows the AGN luminosity function with horizontal and vertical bars presenting the full range in our best-fitting model. Overplotted as a thick black line is the AGN luminosity function from the same observations as in Fig. 2. Our best-fitting model deviates from the observations at $z = 2$ and $z = 0.75$. However, AGN luminosity functions reported in the literature deviate between various surveys. This can be seen when we overplot the AGN luminosity function (dashed red line in Fig. 3) from a large combination of X-ray surveys including XMM and Chandra COSMOS surveys (Miyaji et al. 2015). The discrepancy between Fiore et al. (2012) and Miyaji et al. (2015) is comparable to the discrepancy between our best-fitting model and these observations.

Fig. 4 shows the AGN duty cycle as a function of stellar mass at three redshifts $z = [2.00, 1.25, 0.75]$. 


Figure 2. Black hole mass function at three redshifts \( z = [2.00, 1.25, 0.75] \), for active black holes only, in AGNs with \( \log L_X \ [\text{erg s}^{-1}] \geq 43 \), where \( X = [2–10] \text{ keV} \). Horizontal and vertical bars show the full range of masses in our Monte Carlo realizations. The dotted, blue line shows our BH mass function for all BHs. Overplotted as a thick black line is the active BH mass function for the same luminosity range from observations (HELLAS2XMM) of La Franca et al. (2005, presented in Fiore et al. 2012). Also, the dashed red line shows the local BH mass function for all black holes (Merloni & Heinz 2008). Our best-fitting model is a good match to the observations.

Figure 3. AGN luminosity function at three redshifts \( z = [2.00, 1.25, 0.75] \). Horizontal and vertical bars present the full range in our best-fitting model. Overplotted as a thick black line is the AGN luminosity function from observations (HELLAS2XMM) of La Franca et al. (2005, presented in Fiore et al. 2012). The dashed red line is the AGN luminosity function from a large combination of X-ray surveys including the XMM and Chandra COSMOS surveys (Miyaji et al. 2015). Our best-fitting model is a good match to the observations, although we slightly underpredict the luminosity function towards lower redshifts.
We consider AGNs with log \( L_X \) [erg s\(^{-1}\)] \( \geq 43 \), where \( X = [2–10] \) keV. Horizontal and vertical bars show the full range for the duty cycle in our Monte Carlo realizations. Observations are again from La Franca et al. (2005) and Fiore et al. (2012). Our best-fitting model is a good match to the observations.

Fig. 5 shows the AGN bias factor at three redshifts \( z = [2.0, 1.5, 1.0] \). We have calculated the AGN bias factor using equations (3), (4), and (5) in Cappelluti, Allevato & Finoguenov (2012). From these equations, the AGN bias in a luminosity and redshift range \( \Delta L, \Delta z \) can be written as

\[
\text{bias}(\Delta L, \Delta z) = \frac{\Sigma b_{DMH}(\Delta L, \Delta z)}{N_{AGN}(\Delta L, \Delta z)},
\]

where \( b_{DMH}(\Delta L, \Delta z) \) is the large-scale bias of dark matter haloes that host AGNs in the luminosity and redshift range \( \Delta L, \Delta z \) and \( N_{AGN}(\Delta L, \Delta z) \) is the total number of AGNs hosted by DMHs in the luminosity and redshift range \( \Delta L, \Delta z \). We obtain \( b_{DMH} \) from fig. 11 of Allevato et al. (2011).

Horizontal and vertical bars in Fig. 5 show the full range for AGN bias in our Monte Carlo realizations. The thick black line shows the best-fitting model in Shen (2009). Points are measurements from Croom et al. (2005, green crosses), Porciani & Norberg (2006, green star), Shen et al. (2009, green open square), da Angela et al. (2008, green circles), Myers et al. (2007, green squares) and Allevato et al. (2011, blue triangle). Red squares are from the semi-analytic model of galaxy formation in Gatti et al. (2016). The AGN bias differs substantially from the bias in Shen (2009) at low luminosities. This flattening could be induced by the Monte Carlo approach, as it includes a broad scatter in calculated parameters. Considering the uncertainties in determining the AGN bias, our best-fitting model is a good match to the observations.

Fig. 6 shows the \( M_{BH} - M_{DMH} \) relation in our best-fitting model. Red lines show the full range of the Monte Carlo realizations at redshift \( z = 2 \); blue lines represent the same at \( z = 1 \) and green lines at \( z = 0 \). The thick black line shows the local Ferrarese relation at \( z = 0 \) and it matches our best-fitting model at \( z = 0 \) by default, since we calibrate our model to do exactly that. We find that this match occurs when \( \sigma_{\text{vph}} = 0.77 \sigma_{\text{vir}} \). Our model incorporates no evolution in the \( M_{BH} - \sigma_{\text{vph}} \) relation. It overpredicts the BH mass at high redshift, as BHs grow faster than DMHs. Fig. 6 shows how the \( M_{BH} - M_{DMH} \) relation evolves into the local Ferrarese relation as dark matter haloes grow in mass and ‘catch up’ to the BH growth.

Fig. 7 shows the \( M_{BH} - M_* \) relation in our best-fitting model (top panel). Horizontal and vertical bars show the full range of Monte Carlo realizations at redshift \( z = 0 \). The black line shows the Kormendy & Ho (2013) relation and the dashed green line shows the Merloni & Heinz (2008) relation. Our best-fitting model underpredicts BH masses compared with the Kormendy & Ho (2013) relation. However, the match is better when compared with the Merloni & Heinz (2008) relation. We also find no evolution of scatter in the \( M_{BH} - M_* \) relation in our best-fitting model (bottom panel in Fig. 7). Despite the BH mass being determined by \( \sigma_{\text{vph}} \) via \( \sigma_{\text{vir}} \) and the scatter in \( M_* \), at fixed \( M_{DMH} \), being very small, the resulting \( M_{BH} - M_* \) relation of Fig. 7 is very broad and even significantly below the Kormendy & Ho (2013) relation. This might be in support of the biases in the local scaling relations of BHs and galaxies discussed recently in the literature (Reines & Volonteri 2015; van den Bosch et al. 2015; Greene et al. 2016; Shankar et al. 2016; van den Bosch et al. 2016).

### 3.2 Determining the AGN activity phase

Assuming that major mergers are driving AGN activity in massive galaxies, we have selected simulated mergers of field galaxies in
Figure 5. AGN bias factor at three redshifts $z = [2.0, 1.5, 1.0]$. Horizontal and vertical bars show the full range for AGN bias in our Monte Carlo realizations. The thick black line shows the best-fitting model in Shen (2009). Points are measurements from Croom et al. (2005, green crosses), Porciani & Norberg (2006, green star), Shen et al. (2009, green open square), da Angela et al. (2008, green circles), Myers et al. (2007, green squares) and Allevato et al. (2011, blue triangle). Red squares are from the semi-analytic model of galaxy formation in Gatti et al. (2016). Considering the uncertainties in determining AGN bias, our best-fitting model is a good match to the observations.

Figure 6. $M_{\text{BH}}$–$M_{\text{DMH}}$ relation in our best-fitting model. Red lines show the full range of Monte Carlo realizations at redshift $z = 2$; blue lines represent $z = 1$; green lines $z = 0$. The thick black line shows the local Ferrarese relation at $z = 0$. The figure shows how the $M_{\text{BH}}$–$M_{\text{DMH}}$ relation evolves into the local Ferrarese relation as dark matter haloes grow in mass.

Figure 7. Top panel: $M_{\text{BH}}$–$M_*$ relation in our best-fitting model. Horizontal and vertical bars show the full range of Monte Carlo realizations at redshift $z = 0$. The black line shows the Kormendy & Ho (2013) relation and the dashed green line shows the Merloni & Heinz (2008) relation. Bottom panel: evolution of scatter in the $M_{\text{BH}}$–$M_*$ relation in our best-fitting model (dotted blue line for $z = 2$, dashed red line for $z = 1$ and thick black line for $z = 0$).

Once we find the redshift bin and galaxy-mass bin of the simulated merger, we trace the merging haloes before the merger and the merger remnant after the merger. We determine the initial SMBH mass (before accretion during the AGN phase) and final (‘true’) SMBH mass (after accretion in the AGN phase).
The centres of elliptical galaxies and the scatter in the Kormendy relation at $10^{6.5} - 10^{7} M_{\odot}$ is due to the mass function of the initial BHs in elliptic galaxies peaks at $10^{6.5} - 10^{6.9} M_{\odot}$, depending on the redshift (Fig. 8, left panels). On the other hand, BHs in spiral galaxies, their mass function peaks at $10^{5.5} - 10^{6} M_{\odot}$. After both pre-BHB and BHB phases and after adding the accreted mass to the seed BHs in spiral galaxies, their masses overlap with the masses of BHs in elliptical galaxies. The resulting mass function peaks at $10^{6.5} - 10^{6.9} M_{\odot}$, depending on the redshift (Fig. 8, right panels).

The difference in mass function between the initial and final BHs in Fig. 8 is the accreted mass during the AGN phase in our best-fitting model, where $L_{\text{peak}}$ is calculated from the light-curve model in Shen (2009). The $M_{\text{final}}$ obtained in this manner (the ‘true’ final BH mass) is then compared with $M_{\text{predicted}}$, which is obtained by replacing $L_{\text{peak}}$ with AGN luminosities from the COSMOS survey, $L_{\text{COSMOS}}$.

All galaxies in the specific mass range host AGNs with the probability defined in F14. The probability functions presented in F14 show that galaxies are more likely to host less luminous AGNs. As the observed AGN luminosity increases, the probability of that particular AGN being observed in the COSMOS galaxy decreases. We incorporate COSMOS AGN luminosities into the Shen (2009) model for SMBH growth.

We perform 40 000 Monte Carlo realizations for every $M_{\text{initial}}$ in each $\Delta z$-$\Delta M_*$ interval, through all possible COSMOS AGN luminosities. As a result, we obtain 40 000 predicted BH masses, which we compare with the ‘true’ final BH mass. When $M_{\text{predicted}} \geq M_{\text{final}}$, the COSMOS AGN luminosity is the peak AGN luminosity, corresponding to the AGN luminosity in our best-fitting model. We calculate the percentage of realizations where $M_{\text{predicted}} \geq M_{\text{final}}$ and present it in Fig. 9.

Fig. 9 shows the probability function (occupation fraction) that the observed AGNs are at their peak activity. Nine panels present three redshift ranges and three galaxy log-mass ranges. Thick (black) bars represent probability functions in our model M1. Thin (red) bars represent probability functions in our model M2. Bars

![Figure 8](image_url)
Figure 9. Probability function for the predicted SMBH mass to be at least as large as the true SMBH mass. In other words, the probability that the observed AGN luminosity is large enough to account for the final SMBH mass. Probability functions are split into redshift bins and galaxy mass bins, which correspond to the nomenclature in Bongiorno et al. (2012). Thick (black) histograms represent probability functions in our model M1. Thin (red) histograms represent probability functions in our model M2.

show the full range of Monte Carlo realizations. The probability for peak AGN activity at low redshift \( z = [0.3, 0.8] \) is small in all galaxy mass bins. For \( M_\star = [10^{10.4}, 10^{10.9}] \), all AGNs have probability <20 per cent in model M1 and <30 per cent in model M2. In the same \( \Delta z \sim \Delta M_\star \) interval, there is a probability of <10 per cent of having 90–100 per cent of AGNs in M1 and 80–90 per cent of AGNs in M2. The occupation fraction of AGNs with low probability of peak activity increases towards larger galaxy mass. AGNs hosted by the most massive galaxies (\( M_\star = [10^{10.9}, 11.2] \)) are all in the declining phase of their activity, since the probability drops to <20 per cent in both models (top right panel in Fig. 9). Overall, at low redshift, almost all AGNs are in non-star-forming red sequence galaxies. The increase in the fraction of AGNs with larger probability means more AGNs are in star-forming green valley galaxies. We see this trend as we go from low to high redshift in Fig. 9. At intermediate redshifts (\( z = [0.8, 1.5] \)) the AGN fraction with larger probability increases in both models (middle panels in Fig. 9). As expected, this increase is larger for M2, where \( M_{\text{BH,initial}} \) is larger. Still, most AGNs have low probability of being observed at their peak. AGNs at high redshifts (\( z = [1.5, 2.5] \)) and in the lowest galaxy mass range (\( M_\star = [10^{10.4}, 10^{10.7}] \), bottom left panel in Fig. 9) are dominantly at peak activity, since 30–50 per cent of them in M1 and 55–75 per cent in M2 have >80 per cent probability of being at the peak. Overall, the distribution of occupation fractions shifts towards larger probabilities. Similarly to lower redshifts, the occupation fraction with large probabilities decreases towards more massive galaxies. For \( M_\star = [10^{10.7}, 10^{10.9}] \) (bottom middle panel in Fig. 9), the AGN fraction is evenly distributed. Here, we would expect to see comparable numbers of AGNs in both quiescent and star-forming galaxies. In the largest mass range (panels on the right of Fig. 9), AGNs are predominantly in quiescent galaxies at all redshifts.

The trend that emerges in Fig. 9 is therefore that quiescent galaxies host almost all AGNs at low redshift. As we go towards higher redshift, there are more AGNs in star-forming galaxies and the percentages of AGNs inhabiting quiescent galaxies and star-forming galaxies become comparable. We also see a trend with increasing galaxy mass. At larger galaxy mass, there are more AGNs in quiescent galaxies.

This exact trend is seen in AGNs in the COSMOS survey (figs 12, 13 and 18 of Bongiorno et al. 2012).

4 DISCUSSION AND CONCLUSIONS

We ran a cosmological (130 Mpc box) N-body (dark matter only) simulation from which we located field DMHs at all redshifts. We also followed their evolution while they stay in the field. We found merger events and traced merger progenitors and merger remnants. Through scaling relations, we calculated SMBH masses for progenitors and remnants. In this manner, we obtain the SMBH mass at the centres of DMHs before (initial SMBH) and after (final SMBH) the merger.

We assume that, at the time when the halo merger finishes, a galaxy merger starts. At that time, the newly formed SMBH ignites as an AGN and quickly reaches its peak activity. We focus on two models with different range for the initial BH mass, since BH seeds in spiral galaxies and pre-coalescence growth there are the
source of the largest uncertainty in our modelling. Model M1 has a lower mass range $\sim[10^5-10^6] \, M_\odot$ and a pre-coalescence accretion rate of $10^{-4} \, M_\odot \, yr^{-1}$. Model M2 has a larger initial mass range $\sim[10^5.5-10^6.5] \, M_\odot$ and an accretion rate of $10^{-3} \, M_\odot \, yr^{-1}$.

We determine the ‘true’ final BH mass by using a non-evolving $M_{BH}\sigma_{\text{qph}}$ relation, where $\sigma_{\text{qph}} = 0.77V_{\text{vir}}$. In this manner, the $M_{\text{DMH}}-M_{\text{BH}}$ relation evolves from overestimating BHs masses at high redshift to matching the local Ferrarese relation at $z = 0$.

Our best-fitting model for SMBH growth reproduces the observed AGN luminosity function, SMBH mass function, duty cycle and bias.

Next, we replace peak AGN luminosities in our best-fitting model with COSMOS AGN luminosities from Bongiorno et al. (2012).

For every galaxy hosting an AGN, we determine the redshift and mass and sort them into redshift ranges and mass ranges as in Bongiorno et al. (2012, COSMOS survey). For each mass and redshift, we assign an observed probability function of a galaxy hosting an AGN of a certain luminosity (F14 in Bongiorno et al. 2012). Next, we run 40,000 Monte Carlo realizations in each $\Delta z-\Delta M_*$ interval, where we draw from the observed probability functions, and we assign luminosities to the initial BH. We obtain 40,000 predicted BH masses, which we compare with the ‘true’ final BH mass.

When $M_{\text{predicted}} \geq M_{\text{final}}$, the COSMOS AGN luminosity is the peak AGN luminosity, corresponding to the peak AGN luminosity in our best-fitting model. We calculate the percentage of realizations for which $M_{\text{predicted}}$ is at least as large as $M_{\text{final}}$. A large percentage implies a large probability of AGNs being at their peak activity. A small percentage means that AGNs are not observed at the peak but in the declining phase of their nuclear activity. In this manner, we distinguish ‘peak’ AGNs (recently merger-triggered and hosted by star-forming galaxies; green valley) and ‘faded’ AGNs (merger-triggered a long time ago and now residing in quiescent galaxies; red sequence).

At low redshift range ($z = [0.3, 0.8]$), all observed AGNs are in the declining phase of their nuclear activity, fading away (Fig. 9). The probability of them being at their peak activity is $<10$ per cent for $>90$ per cent of AGNs in the most massive galaxies. The AGN luminosity would have to be very large to account for SMBH growth. However, even if this highest possible luminosity was large enough to produce a final SMBH, it is also the least probable one. Since the entire range of luminosities cannot produce the final SMBH, these luminosities do not correspond to AGN peak activity. The time of maximum nuclear activity, when most of the mass was accreted, has occurred in the past at higher Eddington ratio, when the AGN luminosity was larger and the AGN was most likely hosted by a star-forming galaxy (green valley). The logical conclusion is that the observed luminosities belong to an AGN in the declining phase of its nuclear activity, which places this particular AGN in a quiescent galaxy (red sequence).

Theoretical modelling of AGN populations in hosts of various morphologies, mass ranges and redshifts supports the merger-driven scenario for luminous AGN activity. At the same time, observations are split between the existence of merger features (Koss et al. 2010; Schawinski et al. 2010; Smirnova, Moiseev & Afanasiev 2010; Cotini et al. 2013) and the lack of them (Gabor et al. 2009; Darg et al. 2010; Cisternas et al. 2011; Kocevski et al. 2012; Villforth et al. 2014). Villforth et al. (2014) found no increase in the prevalence of merger signatures with AGN luminosity (in the redshift range $z = [0.5, 0.8]$) and concluded that either major mergers play only a very minor role in the triggering of AGNs in the luminosity range studied ($\log L_X = [41, 44.5]$) or time delays are too long for merger features to remain visible.

Our model shows that the merger-driven scenario is still consistent with the observations, even though there are no merger features in massive galaxies hosting low-redshift AGNs and almost all AGN hosts are quiescent galaxies. How can mergers explain AGN activity in massive galaxies that have no merger features and no star formation to indicate a recent galaxy merger? Since at $z = [0.3, 0.8]$ (Fig. 9) the observed luminosities cannot correspond to AGNs at their peak activity (i.e. cannot produce the final SMBH mass in the simulation), then they must be observed much later in their evolution, long after the merger features can be detected. Our confirmation of Bongiorno et al. (2012)’s results that almost all low-redshift AGNs are in quiescent galaxies is a simple consequence of the drop in galaxy merger rates at $z < 1$. Since galaxy merger rates fall dramatically at low redshift, there are very few recently activated AGNs that would be hosted by star-forming galaxies. Most of the observed AGNs are therefore fading AGNs activated in old mergers that occurred at higher redshifts. Since there are no new galaxy mergers, almost all observed AGNs are in non-star-forming galaxies.

As we go towards higher redshifts, the probability of AGNs being at their peak activity increases. There are more AGNs in star-forming green valley galaxies. At $z = [1.5, 2.5]$, the percentage of AGNs in star-forming galaxies is comparable to the percentage of AGNs in quiescent galaxies. This can be seen in our Fig. 9, bottom panels, and in fig. 18 of Bongiorno et al. (2012). Again, this is a simple consequence of the large merger rate in galaxies at high redshift. The explanation for the comparable number of star-forming and quiescent AGN hosts is that AGNs in star-forming galaxies at high redshift have ‘just’ been triggered by galaxy mergers, while AGNs in quiescent galaxies at the same redshift have been merger-triggered at some time in the past.

Schawinski et al. (2014) had proposed a split of green valley transition into two paths. The current understanding is that late-type galaxies transition slowly from the blue cloud to the red sequence, while hosting low- to intermediate-luminosity AGNs driven by secular processes. Early-type galaxies transition quickly, while hosting high-luminosity AGNs driven by major mergers. In the context of galaxy evolution, our model addresses the evolution of early-type galaxies produced in major mergers of gas-rich disc/spirals. These galaxies correspond to the massive red galaxies in the COSMOS survey (Bongiorno et al. 2012), where they represent the majority of AGN-hosting galaxies. According to our model, AGNs in massive galaxies of the COSMOS survey belong to the rapid transition channel (Schawinski et al. 2014). We find that, right after the merger, AGNs reach their peak activity (green valley phase). This is a short phase ($\sim 100$ Myr) during which star formation is quenched. Then galaxies enter the red-sequence phase, with AGNs in decline (or at the end) of their nuclear activity and low Eddington accretion rate observed in the COSMOS survey. Fig. 2 shows that we are sampling the growth of SMBHs $> 10^7 M_\odot$. For the most part, the final SMBHs are $> 10^8 M_\odot$. This is consistent with Hopkins et al. (2014)’s conclusion that at these masses merger-driven AGN activity dominates.

There are a number of recent articles discussing possible biases in the local scaling relations of BHs and galaxies (Reines & Volonteri 2015; van den Bosch et al. 2015; Greene et al. 2016; Shankar et al. 2016; van den Bosch et al. 2016). In support of this, we find that when $\sigma_{\text{qph}}$ is determined via $V_{\text{vir}}$, the resulting $M_{\text{BH}}-\sigma$ relation of Fig. 7 is very broad and even significantly below the Kormendy & Ho (2013) relation. Shankar et al. (2016) found that the normalization of the $M_{\text{BH}}-\sigma$ relation might be decreased by a factor of 3. Revising the $M_{\text{BH}}-\sigma$ relation in our best-fitting model
leads to smaller final BH masses. This in turn decreases BH mass functions and AGN luminosity functions by a similar factor, but is still consistent with the observations. We have tested how this fact would influence our results and we found that the probability functions in Fig. 9 would shift towards higher probabilities but would not change our results qualitatively.

We conclude that the merger-driven scenario for AGN activity is consistent with observations and that the occupation fractions of the observed AGNs simply follow the evolution of galaxy merger rates. Our model reproduces the observed trend that quiescent (red sequence) galaxies host almost all AGNs at low redshift, due to the dramatic drop in galaxy merger rates at $z < 1$. There are just a few recently activated AGNs in star-forming galaxies. Instead, most AGNs are in their declining nuclear activity hosted by quiescent galaxies.

In fact, one could imagine a scenario where all of the moderate to faint AGNs are secularly driven. However, this does not exclude other mechanisms.

Even though we enforce the criterion on AGN hosts to be $>10^{10.4} \, M_\odot$ in stellar mass, we must have some mixing of populations. We expect that a large majority of AGNs in these galaxies are merger-driven. However, some percentage of AGNs is probably driven by secular processes. That being said, we would like to point out that we are not trying to show that mergers are definitely responsible for AGN activity. We are arguing that the observed AGN activity (at least in $>10^{10.4} \, M_\odot$ galaxies) is consistent with mergers as drivers. However, this does not exclude other mechanisms. In fact, one could imagine a scenario where all of the moderate to faint AGNs are secularly driven.

Two major concerns for our method are the precision in determining the mass of the AGN host galaxy in observations and the detection of low-luminosity AGNs. Both concerns impact the relations between the mass of the host galaxy and the probability functions for AGN incidence.

**ACKNOWLEDGEMENTS**

We thank the anonymous reviewer for tremendous help with the shaping of this article.

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia through project no. 176021, ‘Visible and Invisible Matter in Nearby Galaxies: Theory and Observations’. The authors acknowledge the financial support provided by the European Commission through project BELISSIMA (BELgrade Initiative for Space Science, Instrumentation and Modelling in Astrophysics, call FP7-REGPOT-2010-5, contract no. 256772).

Numerical results were obtained on the PARADOX cluster at the Scientific Computing Laboratory of the Institute of Physics Belgrade, supported in part by the national research project ON171017, funded by the Serbian Ministry of Education, Science and Technological Development.

**REFERENCES**

Allevato V. et al., 2011, ApJ, 736, 99
Barnes J. E., 1988, ApJ, 331, 699
Barnes J. E., Hernquist L. E., 1996, ApJ, 471, 115
Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
Begelman M. C., Rossi E. M., Armitage P. J., 2008, MNRAS, 387, 1649
Behroozi P. S., Wechsler R. H., Wu H. Y., Busha M. T., Klypin A. A., Primack J. R., 2013a, ApJ, 763, 18
Behroozi P. S., Wechsler R. H., Wu H. Y., 2013b, ApJ, 762, 109
Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton Univ. Press, Princeton, NJ
Bomh A., Wisotzki L., GEMS team, 2007, in Metallic N., Shanks T., eds, ASP Conf. Ser. Vol. 379, Cosmic Frontiers. Astron. Soc. Pac., San Francisco, p. 185
Bomh A. et al., 2013, A&A, 549, A46
Bongiorno A. et al., 2012, MNRAS, 427, 3103
Bromm V., Loeb A., 2003, ApJ, 596, 34
Cales S. L. et al., 2013, ApJ, 762, 90
Canalizo G., Stockton A., 2000, AJ, 120, 1750
Canalizo G., Stockton A., 2001, ApJ, 555, 719
Capelo R., Volonteri M., Dotti M., Bollovoy J. M., Mayer L., Governato F., 2015, MNRAS, 447, 2123
Cappelluti N., Allevato V., Finoguenov A., 2012, Adv. Astron., 2012, 853701
Cavaliere A., Vittorini V., 2000, ApJ, 543, 599
Cen R., 2012, ApJ, 755, 28
Ciotti L., Ostriker J. P., Scoppa G., 2007, ApJ, 665, 1038
Ciotti L., Ostriker J. P., Proga D., 2010, ApJ, 717, 708
Cisternas M. et al., 2011, ApJ, 726, 57
Cotini S., Ripamonti E., Caccianiga A., Colpi M., Della Ceca R., Mapelli M., Severgnini P., Segreto A., 2013, MNRAS, 431, 2661
Crocq M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Croom S. M. et al., 2005, MNRAS, 356, 415
Cruton D. J. et al., 2006, MNRAS, 365, 11
da Angela J. et al., 2008, MNRAS, 383, 365
Darg D. W. et al., 2010, MNRAS, 401, 1552
Dressler A., 1980, ApJ, 236, 351
Ellison S. L., Patton D. R., Mendel J. T., Scudder J. M., 2011, MNRAS, 418, 2043
Ellison S. L., Mendel J. T., Patton D. R., Scudder J. M., 2013, MNRAS, 435, 3627
Ferrarese L., 2002, ApJ, 578, 90
Fiore F. et al., 2012, A&A, 537, A16
Gabor J. M. et al., 2009, ApJ, 691, 705
Gaskell C. M., 2009, preprint (arXiv:0908.0328)
Gatti M., Shankar F., Bouillot V., Menci N., Lamastra A., Hirschmann M., Fiore F., 2016, MNRAS, 456, 1073
Gerhard O. E., 1981, MNRAS, 197, 179
Greene J. E. et al., 2016, preprint (arXiv:1606.00018)
Hayward C. C., Torrey P., Springel V., Hernquist L., Vogelsberger M., 2014, MNRAS, 442, 1992
Hernquist L., 1989, Nature, 340, 687
Hopkins P. F., 2012, MNRAS, 420, L8
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Di Matteo T., Springel V., 2006, ApJ, 639, 700
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
