The anti-hierarchical growth of supermassive black holes

Andrea Merloni[★]

Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-85741, Garching, Germany

ABSTRACT

I present a new method to unveil the history of cosmic accretion and the build-up of supermassive black holes (SMBHs) in the nuclei of galaxies, based on observations of the evolving radio and (hard) X-ray luminosity functions of active galactic nuclei (AGN). The fundamental plane of black hole activity discovered by Merloni, Heinz & Di Matteo, which defines a universal correlation among black hole mass (M), 2–10 keV X-ray luminosity and 5-GHz radio luminosity is used as a mass and accretion rate estimator, provided a specific functional form for the dependency of the X-ray luminosity on the dimensionless accretion rate $\dot{m}$ is assumed. I adopt the local black hole mass function (BHMF) as derived from the velocity dispersion (σ) distributions of nearby galaxies coupled with the $M - \sigma$ relation as a boundary condition to integrate backwards in time the continuity equation for the evolution of SMBH, neglecting the role of mergers in shaping up the BHMF. Under the most general assumption that, independently on M, black hole accretion proceeds in a radiatively efficient way above a certain rate, and in a radiatively inefficient way below it, the redshift evolution of the BHMF and the black hole accretion rate (BHAR) function (i.e. the distribution of the Eddington scaled accretion rates for objects of any given mass) are calculated self-consistently. The only tunable parameters are the overall efficiency of extracting gravitational energy from the accreting gas, ϵ, and the critical ratio of the X-ray to Eddington luminosity, $L_{2-10\,\mathrm{keV,cr}}/L_{\mathrm{Edd}} \equiv x_{\mathrm{cr}}$, at which the transition between accretion modes takes place. For fiducial values of these parameters ($\epsilon = 0.1$ and $x_{\mathrm{cr}} = 10^{-3}$), I found that half ($\sim$85 per cent) of the local black hole mass density was accumulated at redshift $z < 1$ ($z < 3$), mostly in radiatively efficient episodes of accretion. The evolution of the BHMF between $z = 0$ and $z \sim 3$ shows clear signs of an anti-hierarchical behaviour: while the majority of the most massive objects ($M \gtrsim 10^9$) were already in place at $z \sim 3$, lower mass ones mainly grew at progressively lower redshift, so that the average black hole mass increases with increasing redshift. In addition, the average accretion rate decreases towards lower redshift. Consequently, sources in the radiatively inefficient regime of accretion only begin to dominate the comoving accretion energy density in the Universe at $z < 1$ (with the exact value of z depending on x_{cr}), while at the peak of the BHAR history, radiatively efficient accretion dominates by almost an order of magnitude. I will discuss the implications of these results for the efficiency of accretion on to SMBH, the lifetimes of quasars and duty cycles, the history of AGN feedback in the form of mechanical energy output and, more generally, for the cosmological models of structure formation in the Universe.

Key words: accretion, accretion discs – black hole physics – galaxies: active – galaxies: evolution – quasars: general.

1 INTRODUCTION

In the last decade, supermassive black holes (SMBHs) in the nuclei of galaxies have been discovered at an ever increasing pace,

and are now believed to reside in most of (perhaps all) the bulges of present-day galaxies (Kormendy & Richstone 1995; Richstone et al. 1998). Moreover, there is evidence of a correlation between the mass of the central black holes and either the mass and luminosity (Kormendy & Richstone 1995; Magorrian et al. 1998; McLure & Dunlop 2002; Marconi & Hunt 2003) or the velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000; Merritt & Ferrarese

[★]E-mail: am@mpa-garching.mpg.de

2001; Tremaine et al. 2002) of its host bulges. This has led to the recognition that the formation and growth of SMBHs and their host galaxies are related processes, and that understanding their evolutionary history can provide fundamental insight into the theories of structure formation in the Universe and into the physical nature of active galactic nuclei (AGN) feedback (Silk & Rees 1998; Fabian 1999; Kauffmann & Haehnelt 2000; Umemura 2001; Granato et al. 2001; Cavaliere & Vittorini 2002; Wyithe & Loeb 2003; Granato et al. 2004).

Within this framework, the most important question that needs to be answered is when and how the mass currently locked up in SMBHs was assembled. In practice, this corresponds to asking what is, at any given redshift z, the number of black holes per unit comoving volume per unit (base 10) logarithm of mass, i.e. the black hole mass function $\phi_M(M, z)$ (hereafter BHMF). In the simplest case of purely accretion driven evolution (i.e. assuming mergers do not play an important role in shaping the BHMF), the simultaneous knowledge of a black hole accretion rate (BHAR) distribution function $\phi_{\dot{m}}(M, z)$, would completely determine the evolutionary solution, as the two distributions must be coupled via a continuity equation (Small & Blandford 1992; Marconi et al. 2004, hereafter M04), to be solved given the appropriate boundary conditions. Moreover, if black holes mainly grew by accretion, a direct link should exist between the quasar (quasi-stellar object, hereafter QSO) and AGN phenomena, signatures of actively accreting phases, and nearby SMBHs tracing directly the entire accumulated mass. In fact, previous comparisons between direct estimates of the local black hole mass density, $\rho_{BH,0} \equiv \rho_{BH}(z = 0) \equiv \int_0^\infty M\phi_M(z = 0)\,dM$, and the total energy density liberated by powerful quasars over the cosmic history, as originally proposed by Soltan (1982), seem to consistently suggest that SMBHs mainly grew while they were active (Salucci et al. 1999; Fabian & Iwasawa 1999; Yu & Tremaine 2002, hereafter YT02; Elvis et al. 2002; Cowie et al. 2003; M04).

From a theoretical point of view, several attempts have been made to link the evolution of the SMBH population to analytic and semi-analytic models of structure formation (Efstathiou & Rees 1988; Haehnelt & Rees 1993; Haehnelt, Natarajan & Rees 1998; Haiman & Loeb 1998; Cattaneo, Haehnelt & Rees 1999; Kauffmann & Haehnelt 2000; Monaco & Salucci 2000; Cavaliere & Vittorini 2000; Volonteri, Haardt & Madau 2003; Wyithe & Loeb 2003; Hatziminaoglou et al. 2003; Bromley, Somerville & Fabian 2004; Mahmood, Devriendt & Silk 2004). A different approach has been followed recently by Di Matteo et al. (2003), who made use of large cosmological hydrodynamical simulations to predict the evolutionary history of the growth and activity of SMBHs.

The path from cosmological cold dark matter (CDM) structure formation models (and/or simulations) to the predicted AGN evolving luminosity functions, however, is dotted with uncertainties regarding a number of important physical processes. In general, semi-analytic modellers need to make specific assumptions (and introduce parameters) to describe gas cooling in dark matter haloes, the fuelling of central nuclear black holes, the physics of merger events and the nature of various feedback mechanisms (from star formation and/or from AGN). To constrain some of these parameters, all these models need to be tested against observationally determined signatures of the growth history of SMBHs. The usual test-benches are: the evolution of the AGN luminosity functions (in any specific band), the local BHMF, as derived from the $M - \sigma$ relation, the slope and intercept of the $M - \sigma$ relation itself and the spectrum and intensity of the X-ray background light, known to be produced by the sum of individual AGN.

Here I would like to propose an alternative approach which is capable of providing a self-consistent evolutionary picture for SMBHs. Such an approach is largely based on *observed* data and only on a minimal number of theoretical assumptions (and parameters). These assumptions are only needed to describe the physics of the innermost accretion process, through which black holes shine, and not the more complex interplay between growing SMBHs and their galactic environment. This is an advantage for two reasons: first of all, the physics of black hole accretion is reasonably well understood both theoretically and phenomenologically; secondly, the physical properties of the innermost part of an accretion flow, where the dynamics is almost completely dominated by the strong gravitational field of the central black holes, should not depend on cosmology and redshift.

The approach followed here is similar in spirit to those of Small & Blandford (1992), Salucci et al. (1999), YT02 and M04. More specifically, I will make use of the local BHMF and of the evolution of AGN luminosity functions to constrain the evolutionary history of SMBHs. The main novelty of the present work is the realization that simultaneous radio and (hard) X-ray observations of accreting black holes can provide tight constraints on *both* the mass and the accretion rate of an active black hole, through the so-called 'fundamental plane' relationship for active black holes (Merloni et al. 2003, hereafter MHD03). From this, the history of the growth of an SMBH can be followed in detail up to the redshift at which reliable X-ray and radio luminosity functions can be obtained.

The structure of the paper is as follows. In Section 2 I will describe how knowledge of the X-ray and radio luminosity functions of local AGN and the local SMBH mass function can be used to obtain a complete census of the local SMBH population and activity distribution in the form of an accretion rate distribution function. A necessary ingredient for performing this calculation is the functional form that relates the observed X-ray luminosity of an accreting black hole to its mass and accretion rate. In Section 3 I will describe how it is possible to use our theoretical and phenomenological knowledge of the different modes of accretion on to a black hole to obtain such a relation. In Section 4, the method to calculate the redshift evolution of both black hole mass and accretion rate functions is described, and its basic assumptions clearly spelled out. Section 5 is then devoted to the analysis of the most important results of the calculations and of the main properties of the evolving SMBH population between $z = 0$ and $z \sim 3$. A more general discussion of the implication of these results is presented in Section 6. Finally, I draw my conclusions in Section 7.

Throughout this paper, we adopt a background cosmological model in accordance with the Wilkinson Microwave Anisotropy Probe (WMAP) experiment. The model has zero spatial curvature, a cosmological constant, $\Omega_\Lambda = 0.71$ a Hubble constant $H_0 = 70\,\mathrm{km\,s^{-1}}$, dominated by cold dark matter with $\Omega_m = 0.29$ and $\Omega_b = 0.047$ (Spergel et al. 2003).

2 THE IMPORTANCE OF ESTIMATING THE CONDITIONAL RADIO/X-RAY ACTIVE GALACTIC NUCLEI LUMINOSITY FUNCTION

Large optical surveys carried out in recent years (Boyle et al. 2000; Fan et al. 2001; Wolf et al. 2003) probe the evolution of the QSO luminosity function up to high redshift, and all agree in establishing a strong rise in their activity from the local universe up to redshift $z \sim 2$ and a decline above $z \sim 3$. However, inferring the properties of the entire class of accreting SMBHs from samples selected in a single waveband can be misleading, in particular in the case of

the optical band, where obscuration effects can be significant. Indeed, Barger et al. (2003) have clearly shown that optically selected broad-line AGN and QSO form only about one third of the X-ray background: hard X-ray selected samples, therefore, provide a more direct probe of SMBH activity (see also Cattaneo & Bernardi 2003). Recent works by Hosokawa (2004) and M04 have also demonstrated that the redshift evolution of the hard X-ray luminosity function (HXLF) better describes the growth history of accreting SMBHs. Discrepancies with the results obtained from optically selected QSO luminosity functions for the average accretion efficiency and the local black hole mass density can be understood by taking into account the luminosity dependence of both obscuration and bolometric corrections in the different bands (see e.g. Ueda et al. 2003; Cattaneo & Bernardi 2003; Hosokawa 2004, M04).

Similarly, hard X-ray emission reveals fundamental properties of an accreting black hole as follows. In a recent paper (MHD03) it has been shown that if we define the instantaneous state of activity of a black hole of mass M (in units of solar masses), by the radio (at 5 GHz, L_R) and hard (in the 2–10 keV band, L_X) X-ray luminosity of its compact core, and represent such an object as a point in the three-dimensional space ($\log L_R$, $\log L_X$, $\log M$), all black holes (either of stellar mass or supermassive) will lie preferentially on a plane (the 'fundamental plane' of black hole activity), described by the following equation:

$$\log L_R = \left(0.60^{+0.11}_{-0.11}\right)\log L_X + \left(0.78^{+0.11}_{-0.09}\right)\log M + 7.33^{+4.05}_{-4.07}. \quad (1)$$

Equation (1) can be inverted to relate BH masses to observed nuclear radio and X-ray luminosities, i.e.

$$\log M \simeq 1.28(\log L_R - 0.60\log L_X) - 9.34 \pm 1.06$$
$$\equiv g(\log L_R, \log L_X). \quad (2)$$

This is an entirely empirical relation, and as such is independent on any accretion (or jet) physical model. It shows, however, that disc and jet emission from active black holes of any mass, from galactic X-ray binary sources to the most powerful quasars, are physically and observationally correlated phenomena. Moreover, as the fundamental plane relationship is obeyed by the intrinsic hard X-ray luminosities, it is basically unaffected by absorption, and therefore largely independent on the validity of any specific unification scheme for AGN.

One of the consequences of this relationship is that, in an ideal case, the conditional radio/X-ray luminosity function of active black holes, i.e. the number of sources per unit comoving volume per unit logarithm of radio and X-ray luminosity, $\Psi_C(L_R, L_X)$, could be used to reconstruct the mass function of the underlying black hole population. The importance of such a possibility should not be underestimated. For example, it could provide an alternative way to study the demography of the SMBH population (at any redshift) to be compared with what was obtained from the $M - \sigma$ relation [which the Sloan Digital Sky Survey (SDSS) will largely contribute to, see e.g. McLure & Dunlop 2004], or with any other analysis based on different mass estimators (see e.g. Vestergaard 2004). Alternatively, such comparisons can be used to test the redshift evolution of the fundamental plane relationship itself.

Although the future goal for solving the problem at hand should therefore be identified with the study of large multiwavelength (X-ray and radio in particular) SMBH samples, nevertheless I will argue here that it is still possible to make some progress with the currently available pieces of information. In fact, the lack of the exact knowledge of the conditional radio/X-ray AGN luminosity function can be (at least partly) superseded, given the two separate radio, $\phi_R(L_R, z)$, and X-ray, $\phi_X(L_X, z)$, luminosity functions

at redshift z, and an independent estimate of the BHMF, $\phi_M(M, z)$ at the same redshift. By taking into account the fundamental plane relationship, we have that the conditional luminosity function Ψ_C has to satisfy the following integral constraints:

$$\phi_X(L_X)\,d\log L_X = \int_{L_{R,min}}^{\infty} \Psi_C(L_X, L_R)\,d\log L_R \quad (3)$$

$$\phi_R(L_R)\,d\log L_R = \int_{L_{X,min}}^{\infty} \Psi_C(L_X, L_R)\,d\log L_X \quad (4)$$

$$\phi_M(M)\,d\log M =$$
$$\iint_{\log M < g < \log M + d\log M} \Psi_C(L_X, L_R)\,d\log L_R\,d\log L_X, \quad (5)$$

where $g(L_R, L_X)$ is defined in equation (2). In the above formulae, the lower end of the X-ray luminosity function and RLF, $L_{X,min}$ and $L_{R,min}$, should be chosen in such a way as to give the same total number of objects with mass larger than a certain minimal value M_{min}, as computed by integrating the BHMF.

In the following, I will make the assumption that any function $\Psi_C(L_X, L_R)$ that satisfies equations (3), (4) and (5) can be regarded as the true conditional radio/X-ray luminosity function. To begin with, I will show in the next section how to use the above formalism to derive information about the state of activity of the local SMBH population, in the specific form of its accretion rate distribution function.

2.1 The census of the local population

Using the integral constraints provided by equations (3)–(5), it is possible to deduce a conditional radio/X-ray AGN luminosity function at zero redshift. To do that, a specific choice of the observed luminosity (and mass) functions needs to be carried out. I will adopt the following.

(i) For the HXLF I use the recently estimated one of Ueda et al. (2003). This is arguably the most complete luminosity function in the 2–10 keV spectral band, spanning the luminosity range of $10^{41.5}$–$10^{46.5}$ erg s^{-1}, corrected for absorption.

(ii) The 5-GHz RLF of AGN is here derived from the lower frequency one of Willott et al. (2001), obtained from the 3CRR, 6CE and 7CRS complete samples, assuming for the radio spectral index a constant value $\alpha_R = 0.7$ to rescale the luminosities. Although other determinations of the local AGN RLF at higher frequencies are available (see e.g. Sadler et al. 2002), the Willott et al. one is probably the most accurate to date up to relatively high redshifts, and this will be instrumental in studying the growth history of the SMBH population (see Section 4).

(iii) The local BHMF is estimated following Aller & Richstone (2002). They derived it from the local luminosity functions of galaxies of different morphologies given in Marzke et al. (1994), together with the empirical relationships among total luminosity, bulge luminosity and velocity dispersion. However, differing from Aller & Richstone (2002), we do take into account the effects of the scatter in the $M - \sigma$ relation,[1] in the same way as described in Yu & Lu

---

[1] I will assume throughout the paper that the black hole mass-velocity dispersion relation is given by $\log M_{BH} = 8.18 + 4.02\log(\sigma/200\,km\,s^{-1})$, as discussed in Tremaine et al. (2002).

(2004). Although the overall effect of considering such a scatter affects the value of the total black hole mass density only marginally (a dispersion of 0.27 dex results in an increase in the local black hole mass density of a factor of 1.2, see YT02), nevertheless, the shape of the local BHMF is strongly affected, as demonstrated by M04.

Starting with an initial guess for Ψ_C, we proceed via successive iterations, minimizing the differences between the projections of the conditional luminosity function on to the X-ray and radio luminosity axes and the observed luminosity functions, until we obtain a conditional LF that simultaneously satisfies the integral constraints given by equations (3)–(5).

Once such an estimate of the conditional luminosity function is found, it is possible to derive the local distribution of the second fundamental physical parameter that characterizes any active black hole: i.e. its accretion rate in units of Eddington luminosity

$$\dot{m} \equiv \epsilon \dot{M} c^2 / L_{\text{Edd}} \tag{6}$$

(ϵ is the accretion efficiency, see Section 3.1). Such an inversion, however, is model dependent, as it depends on the choice of which spectral energy distribution should correspond to any specific couple of fundamental parameters M and $\dot{m}$. In practice, it corresponds to the choice of the accretion mode of a SMBH of given mass and 2–10 keV luminosity. This can be achieved by choosing a specific functional form $L_X = L_X(M, \dot{m})$, as I will describe in more detail in the next section.

The population of local black holes is dominated by sources shining, in the X-ray band, below 10^{-3} of the Eddington luminosity (assuming a 10 per cent efficiency, see below). This is in agreement with the average value found using the X-ray luminosity function of type 1 Seyfert galaxies by Page (2001). A more detailed view of the local accreting SMBH population can be obtained by studying the mean dimensionless accretion rate as a function of black hole mass, defined as

$$\langle \dot{m}(M) \rangle = \frac{\int_M^{M+dM} \dot{m}(M') \phi_M(M') \, dM'}{\int_M^{M+dM} \phi_M(M') \, dM'}. \tag{7}$$

In Fig. 1 such a quantity is plotted versus the SMBH mass at $z = 0.1$. It is clear that the mean accretion rate is a strong function of black hole mass, with small black holes accreting at a higher rate. Similarly, if we define the instantaneous growth rate of a SMBH of mass M as $M/\langle \dot{M}(M) \rangle$, we see in Fig. 2 that the growth time is very large (about one order of magnitude larger than the Hubble time) for the more massive holes, a result confirmed by the SDSS optical study of 23 000 local AGN (Heckman et al. 2004). This implies that the very high mass SMBH must have formed at significantly higher redshift, as we will see in detail in Section 5, where the redshift evolution of the SMBH population will be calculated.

Before entering into a detailed description of the results on the redshift evolution of the SMBH population obtained using the local distributions as a boundary condition, I will discuss in the next section the main theoretical assumptions underlying such a calculation.

3 THEORETICAL ASSUMPTIONS

3.1 Accretion efficiency

The accretion efficiency ϵ, appearing in equation (6) represents the efficiency with which gravitational energy of the matter infalling on to the black hole can be extracted, regardless of it being trans-
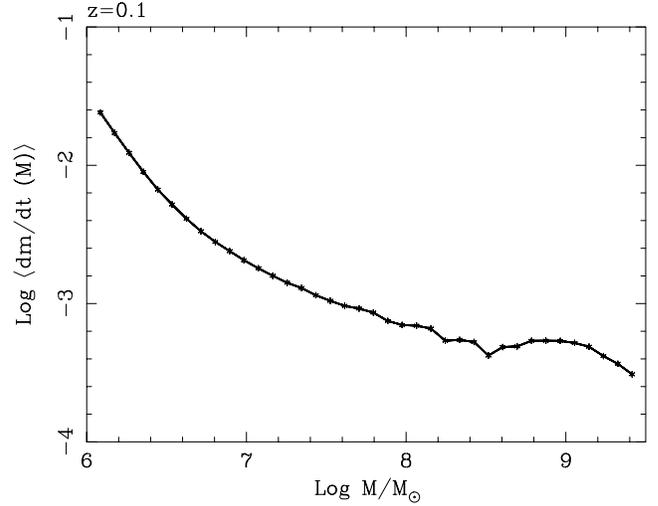

Figure 1. The mean accretion rate (in units of Eddington) $\langle \dot{m}(M) \rangle$ as a function of black hole mass for local ($z = 0.1$) accreting black holes.
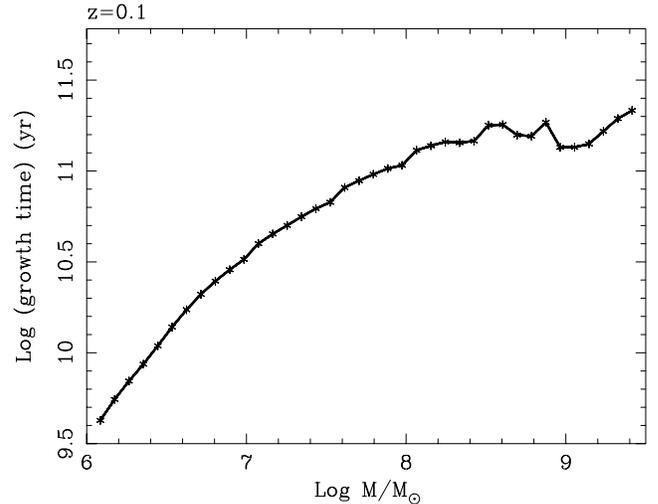

Figure 2. The instantaneous growth rate of local SMBHs, defined as $M/\langle \dot{M}(M) \rangle$ is plotted as a function of black hole mass M.

formed into radiation. It is, therefore, an upper limit to the radiative efficiency, ϵ_{rad}, and a function of just the inner boundary condition of the accretion flow. The closest the innermost stable circular orbit (ISCO) of the accreting gas is to the event horizon, the higher the accretion efficiency. In the classical general relativistic case of test particles, the position of the ISCO depends on the dimensionless angular momentum of the black holes, and the corresponding efficiency varies between $\epsilon \simeq 0.06$ for non-rotating holes and $\epsilon \simeq 0.42$ for maximally rotating Kerr black holes (Novikov & Thorne 1973). Recently, both YT02, comparing the local BHMF with the black hole mass density accreted during luminous QSO phases, and Elvis, Risaliti & Zamorani (2002), comparing the X-ray background intensity with the local SMBH mass density, came to the conclusion that most SMBHs must be rapidly spinning, i.e. they must have accreted at an average radiative efficiency higher than the canonical value of 10 per cent, so that $\epsilon \gtrsim \epsilon_{\text{rad}} \gtrsim 0.1$. Such a conclusion, however, has been questioned by M04 on the basis of a revised estimate of the local SMBH mass function and of the HXLF of AGN.

In what follows, I will assume that the accretion efficiency is constant, regardless of the nature of the specific accretion mode of each SMBH. The *radiative* efficiency, instead, will be a function of the accretion rate, expressed though the dependence of the observable 2–10 keV luminosity on $\dot{m}$ and of the bolometric correction, as described below.

3.2 Accretion mode transition and bolometric corrections

The relevance of the different (theoretical) accretion modes for the various AGN populations is still a matter of open debate. Here I will follow the approach of MHD03 and try to maximize the amount of information on the issue by comparing active black holes of different masses.

Galactic (stellar mass) black holes in X-ray binaries, either of a transient or persistent nature, are commonly observed undergoing so-called transitions, i.e. dramatic changes in their spectral and variability properties (I will adopt here the terminology of McClintock & Remillard 2004, which the reader is referred to for a recent review). There are at least three well-defined spectral states. In the low/hard state the emission is dominated by a hard X-ray power law with an exponential cut-off at about few 100 keV. The spectrum of the thermal dominant (or high/soft) state, instead, is dominated by a thermal component likely originated in a standard Shakura & Sunyaev (1973) accretion disc, while in the steep power-law (or very high) state, usually associated with the highest flux level of a source, both a thermal and a steep power-law component substantially contribute to the spectrum. Maccarone (2003) has shown that the hard-to-soft transition in these systems generally occurs at X-ray luminosities (in the 2–10 keV band) of about $x_{\rm cr} \simeq 5 \times 10^{-3}$. This transition is also accompanied by a 'quenching' of the steady radio emission observed in the low/hard state.

Given the many similarities between the high-energy spectra of galactic black holes and AGN, a similar phenomenology has long been searched for in active SMBHs. The reader is referred to MHD03 for a thorough discussion on the scale-invariant properties of black holes coupled accretion/jet systems, and on how to constrain theoretical accretion models for the different classes of objects on the basis of the observed fundamental plane correlation coefficients. There we showed that, for black holes of any mass characterized by $x \equiv L_{\rm X}/L_{\rm Edd} \lesssim 0.01$, the fundamental plane relation is consistent with the most general theoretical relation between radio emission, mass and accretion rate expected from synchrotron emitting jets (regardless of their detailed geometrical and kinematical properties), provided that the X-ray emitting flow is radiatively inefficient (RIAF; for a recent review about RIAF, see Narayan 2002 or Quataert 2003). In this case, $L_{\rm X} \propto \dot{m}^{2.3}$, and the radio luminosity satisfies: $L_{\rm R} \propto \dot{m}^{1.38} M^{1.38} = \dot{M}^{1.38}$, i.e. $L_{\rm R}$ scales with only the *physical accretion rate*.

On the other hand, the most luminous sources, as those falling into the standard definition of quasars and broad-lined AGN, must be accreting at an higher rate, close to the Eddington one. In addition, their spectral energy distribution, usually dominated by the so-called Big Blue Bump [quasi-thermal ultraviolet (UV) emission most likely from an optically thick standard accretion disc, see e.g. Malkan 1983; Laor 1990], indicates that above a certain critical X-ray to Eddington ratio $x_{\rm cr}$, an accretion mode transition should take place to what is usually described as a standard, geometrically thin and optically thick accretion disc (Shakura & Sunyaev 1973). More recently, Maccarone, Gallo & Fender (2003), analysing the same AGN sample of MHD03, have found evidence of a

connection between the radio-quiet AGN and X-ray binaries in the thermal dominant state. In addition, the transition luminosity, $x_{\rm cr}$, has been found to be consistent with the hard-to-soft transition of galactic black holes.

The scaling of the the hard X-ray luminosity with the accretion rate in this high-$\dot{m}$ regime, $L_{\rm X} \propto \dot{m}^q$, is not straightforwardly predicted by standard accretion disc theory, as the origin of the hard X-ray emission itself is not self-consistently predicted by this theory (but see, for example, Merloni 2003). On the other hand, observational studies of the spectral energy distributions of a large number of QSO and AGN and comparisons of X-ray and optical luminosity functions of AGN can be used to put constraints on the value of q (Ueda et al. 2003).

In light of these facts, I will adopt here the simplest possible functional form for the $L_{\rm X}/L_{\rm Edd}$ versus $\dot{m}$ function of a broken power law, bridging the low accretion rate (RIAF) regime and the high accretion rate one. For radiatively efficient sources, we should assume, by definition, that the bolometric luminosity is simply proportional to the accretion rate, $L_{\rm bol} \propto \dot{m}$. Then, using the fitting formula for the X-ray to bolometric luminosity ratio of M04, I obtain $x \propto \dot{m}^{0.76}$. Summarizing, the universal accretion mode function adopted here has the following scalings:

$$x \propto \begin{cases} \dot{m}^{2.3} & x \lesssim x_{\rm cr} \\ \dot{m}^{0.76} & x > x_{\rm cr} \end{cases} \tag{8}$$

The overall normalization is found by imposing continuity at $x_{\rm cr}$ and that in the radiatively efficient regime $\log x/(0.76 \log \dot{m}) = -1.5$, such that AGN at the peak of the mass distribution have X-ray to optical ratios consistent with observational data (see e.g. Vignali, Brandt & Schneider 2003).

4 THE REDSHIFT EVOLUTION OF THE BLACK HOLE MASS AND ACCRETION RATE FUNCTIONS

In Section 2.1 I have briefly described the general method by which to derive the mass and accretion rate distribution functions of the local SMBHs given the luminosity functions of AGN in both radio and X-ray bands (down to a sufficiently low luminosity in each band in order to match the integrated number densities). From these, the redshift evolution of the SMBH population can be computed integrating *backwards* the continuity equation that describes SMBH evolution driven by accretion only (Small & Blandford 1992; Steed & Weinberg 2004; Hosokawa 2004), i.e.

$$\frac{\partial \phi_{\rm M}(M, t)}{\partial t} + \frac{\partial [\phi_{\rm M}(M, t) \cdot \langle \dot{M}(M, t) \rangle]}{\partial M} = 0, \tag{9}$$

where the mean accretion rate, $\langle \dot{M} \rangle$, as a function of black hole mass and time can be calculated directly from the accretion rate distribution function at time t. By setting the right-hand side of equation (9) to zero, I have implicitly assumed that mergers are unimportant for the growth history of the black hole. A thorough examination of the possible consequences for SMBH growth of the inclusion of additional merger or direct formation terms in equation (9) can be found in YT02, Hosokawa (2004) or Menou & Haiman (2004).

In practice, starting from the BHMF and the accretion rate function at a given redshift z, it is possible to derive the new BHMF at redshift $z + {\rm d}z$, $\phi_{\rm M}(M, z + {\rm d}z)$, by just subtracting the mass accreted in the time interval ${\rm d}t = {\rm d}z({\rm d}t/{\rm d}z)$ calculated according to the accretion rate function of redshift z. This new BHMF can then be used together with the radio and X-ray luminosity functions at the

same redshift, $\phi_R(L_R, z + dz)$ and $\phi_X(L_X, z + dz)$, to obtain the new conditional luminosity function $\Psi_C(L_R, L_X, z + dz)$, and therefore the new accretion rate function, and so on. Thus, the local BHMF, as determined independently from the galaxy velocity dispersion distribution and the $M - \sigma$ relation, and the local accretion rate distribution function, as derived from the conditional radio/X-ray LF and a specific accretion modes scenario, i.e. the function $x(\dot{m})$, can be used together as a boundary condition to integrate equation (9) up to the redshift where the HXLF and the RLF of AGN can be reliably estimated. At each redshift, I discard all SMBH whose mass has decreased below $M_{min} = 10^6 \, M_\odot$, so that nothing can be said, within such a scheme, about the formation of the SMBH seeds. Furthermore, at every redshift, the values of $L_{X,min}$ and $L_{R,min}$ (see equations 3 and 4) are increased to take into account the loss of SMBH below our threshold. In this way, at every redshift the number of objects above a certain minimal mass is always equal to the number of radio and X-ray sources above the corresponding limiting luminosities.

The final result will depend on the following assumptions.

(i) The X-ray luminosity function and the RLF of AGN, if extrapolated down to a sufficiently low luminosity, do indeed describe the same class of objects (SMBH), and this class of objects is assumed to be described by the mass function ϕ_M. Apart from physical arguments invoking the same inner engine for radio and X-ray selected AGN, there are indeed strong similarities between the cosmic evolution of radio sources and of X-ray and optically selected AGN (Dunlop 1998; Willott et al. 2001), a further hint of the correctness of this assumption.

(ii) The fundamental plane of black hole activity is the same at all redshifts, and the correlation coefficients are the same as those observed at zero redshift. As remarked in the Introduction, this is a natural consequence of the fact that the fundamental plane is a relationship between physical quantities in the innermost region of the coupled accretion flow-jet system. There, the strong black hole gravitational field dominates, and cosmological evolution is negligible. However, it should be noted here that the radio emission from steep spectrum AGN, whose RLF we adopt here, may be sensitive to properties of the interstellar and intergalactic medium which indeed should depend on redshift. Ideally, the redshift evolution of the luminosity function of the AGN radio cores should be used, as the fundamental plane (equation 1) is a relationship between core luminosities only.

(iii) The evolution with redshift of both ϕ_R and ϕ_X is known. This is certainly more accurate for the HXLF, and in the following we adopt the luminosity dependent density evolution (LDDE) model of the 2–10 keV luminosity function derived by Ueda et al. (2003). The analytic approximation to the HXLF is described in Appendix A.

On the other hand, the high-redshift evolution of the RLF is much less certain. The best estimate to date is probably that by Willott et al. (2001). This is based on three redshift surveys of flux-limited samples of steep spectrum sources selected at low frequencies. By selecting only steep spectrum sources the authors made sure that the effect of strongly beamed sources (which have typically flat spectra) were minimized, especially for the highest redshift sources. However, the fundamental plane relationship was determined by including both flat and steep spectrum sources (but excluding beamed objects, see MHD3), and the selection was made at higher frequencies. Despite this potential source of uncertainty, here I will adopt the Willott et al. (2001) parametrization, assuming a uniform radio spectral index $\alpha_R = 0.7$ to rescale the fluxes. Previous studies

of redshift evolution of the RLF of flat and steep spectrum sources (Dunlop & Peacock 1990) have shown that although steep spectrum sources dominate number counts, the two populations display similar redshift evolution. The analytical approximation to the RLF of Willott et al. (2001) used here can be found in Appendix B.

(iv) The function $L_X/L_{Edd} = x(\dot{m})$ does not depend on the black hole mass and is expressible as a broken power law: at low accretion rates black holes accrete in a radiatively inefficient way, and $x \propto \dot{m}^{2.3}$, while at high accretion rates black holes are radiatively efficient, with the bulk of the emission being radiated in the optical/UV bands, so that $x \propto \dot{m}^{0.76}$ (see Section 3).

The whole history of the growth of SMBHs can then be reconstructed from the evolution of X-ray and radio AGN luminosity functions with just three free parameters: the accretion efficiency ϵ; the value of the critical ratio x_{cr} that separates the radiatively inefficient and efficient regimes; and the corresponding critical accretion rate or, equivalently, an X-ray bolometric correction.

5 RESULTS

The family of all possible outcomes of the calculation outlined in Section 4 is obtained by varying parameters in the relatively narrow range of physically and phenomenologically realistic values: $0.06 < \epsilon < 0.42$ and $10^{-4} \lesssim x_{cr} \lesssim 10^{-2}$. However, the uncertainties on the observed luminosity functions, the high-redshift radio one in particular, do not allow to put tight constraints on these parameters. More interesting, instead, is the overall trend emerging from this calculation for the cosmic evolution of the population of the SMBH and of its activity level. For this reason, here I discuss in detail the results of a calculation performed assuming $\epsilon = 0.1$ and $x_{cr} = 10^{-3}$, leaving a discussion of the consequences of a different choice of parameters to Section 5.1.

Fig. 3 shows, in the upper panel, the redshift evolution of the comoving black hole mass density from $z = 3.5$ to $z = 0$. The local black hole mass density is $\rho_{BH}(z = 0) = 2.6 \times 10^5 \, M_\odot \, Mpc^{-3}$, a value consistent with that obtained by YT02, as already discussed
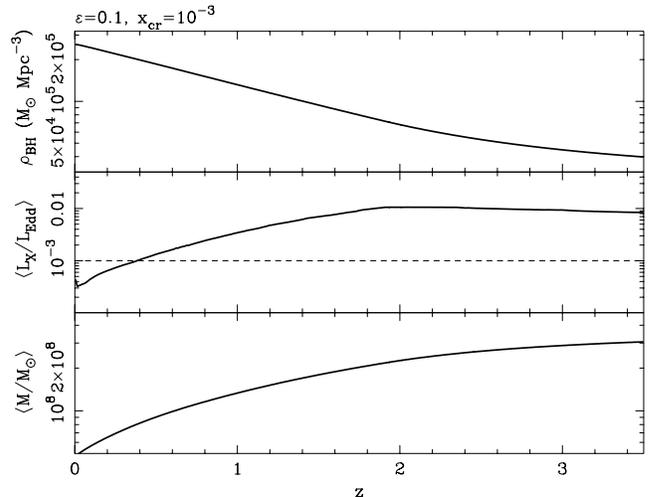

Figure 3. Redshift evolution of the comoving black hole mass density ρ_{BH} (top panel), of the average X-ray to Eddington ratio, $\langle L_X/L_{Edd} \rangle$ (middle panel) and of the average SMBH mass, in units of solar masses (bottom panel). The dashed horizontal line on the middle panel marks the adopted value of the critical accretion rate, x_{cr}, separating RIAF accretion from radiatively efficient one: most of the SMBH growth, therefore, took place during episodes of radiatively efficient accretion (see also Fig. 5).

by Aller & Richstone (2002), while the SMBH mass density at redshift $z = 3$ is about six times lower: $\rho_{BH}(z = 3) = 4.5 \times 10^4 \, M_\odot$ Mpc^{-3}. Half of the total black hole mass density was accumulated at redshift $z < 1$. This is consistent with the most recent results from X-ray background (XRB) studies. With the deep *Chandra* and *XMM* surveys revealing the redshift distribution of the obscured sources that most contribute to the XRB (Alexander et al. 2001; Barger et al. 2002; Mainieri et al. 2002; Rosati et al. 2002; Hasinger 2003; Fiore et al. 2003), the newest synthesis models seem to suggest that indeed a substantial fraction of the locally measured mass density of SMBH (maybe up to 50 per cent) was accumulated in (obscured) [2] low-mass AGN (with $M \lesssim 10^8 \, M_\odot$) at $z < 1$ (see e.g. Gandhi & Fabian 2003 or Fabian 2004 and references therein). This is also consistent with the general picture for the BHMF evolution discussed below.

The middle panel of Fig. 3 shows the evolution of the average X-ray to Eddington rate, defined as

$$\left\langle \frac{L_x}{L_{Edd}}(z) \right\rangle \equiv \langle x(z) \rangle \equiv \frac{\int_{x_{min}}^{\infty} x \phi_x(z) \, dx}{\int_{x_{min}}^{\infty} \phi_x(z) \, dx}. \tag{10}$$

Here $\phi_x(z)$ is the X-ray to Eddington ratio function, representing the number of sources per unit comoving volume per unit logarithm of the X-ray to Eddington ratio. It is simply related to the accretion rate function thorough the monotonic function of equation (8): $\phi_x(z) = \phi_{\dot{m}}(z) \, d\dot{m}/dx$. In the same panel, the dashed horizontal line marks the value of x_{cr}. The average accretion rate increases by almost an order of magnitude from zero redshift until $z \sim 2$, where the luminosity density of the hard X-ray selected sources peaks, and then levels off. SMBHs were more active in the past, in the sense that their average dimensionless accretion rate was higher at higher redshift. A similar conclusion was drawn by Small & Blandford (1992), who adopted a phenomenological approach close to the one followed here, and by Haiman & Menou (2002) and Menci et al. (2003) in the framework of semi-analytic models for structure formation in cold dark matter universes (see discussion in Section 6).

The bottom panel of Fig. 3 shows the average black hole mass, i.e.

$$\langle M(z) \rangle \equiv \frac{\int_{M_{min}}^{\infty} M \phi_M(z) \, dM}{\int_{M_{min}}^{\infty} \phi_M(z) \, dM}, \tag{11}$$

where at every redshift we have taken $M_{min} = 10^6 \, M_\odot$. The anti-hierarchical (Granato et al. 2001, 2004, M04) nature of the growth of SMBHs is summarized by this plot, showing the increase of the average black hole mass with increasing redshift.

More specifically, the evolution of the *shape* of the BHMF, which in turn determines the evolution of $\langle M \rangle$, is shown in Fig. 4. As opposed to the standard picture of hierarchical mass build up of dark matter haloes in CDM cosmologies, SMBHs growing by accretion between $z \sim 3$ and now have a mass function which is more and more dominated by largest mass objects at higher and higher redshift, at least up to the limit where we can trust the evolution of our luminosity functions. As is indeed emerging from the study of the QSO population of the SDSS (Vestergaard 2004; McLure & Dunlop 2004), most of the more massive black holes ($M \gtrsim 10^9$) were already in place at $z \sim 3$.

---

[2] It is worth emphasizing that both the fundamental plane relationship and the HXLF used here rely on absorption corrected 2–10 keV luminosities, and are therefore unaffected by Compton thin absorption. Therefore, throughout the paper no distinction is made (and is possible) between obscured and unobscured sources.
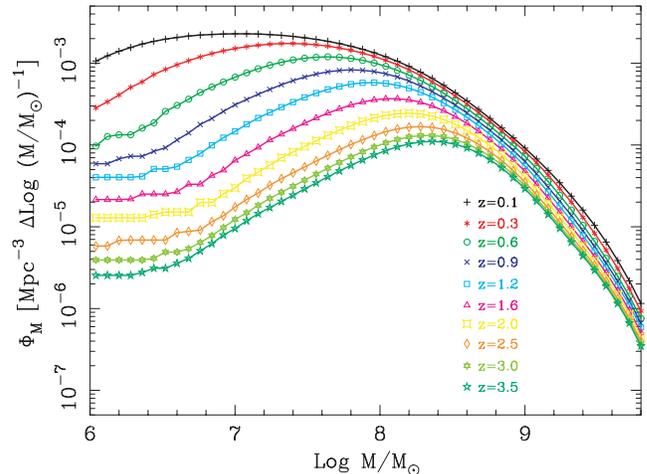
Figure 4. Redshift evolution of the SMBH mass function (BHMF), from redshift $z = 3.5$ till redshift $z = 0.1$. Different colours and symbols correspond to different redshift bins.
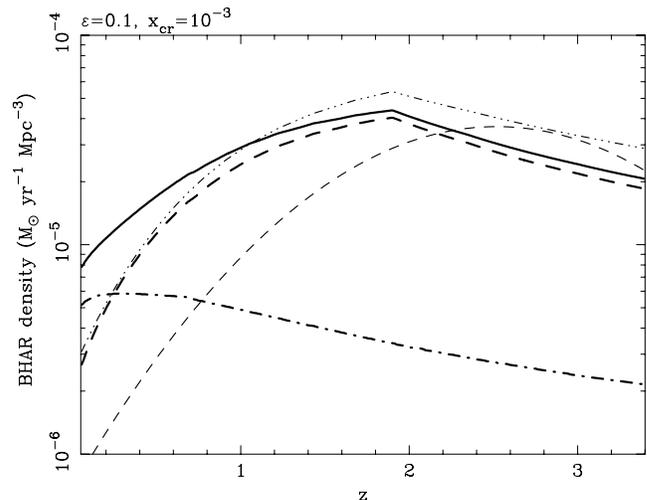

Figure 5. Comparison of the evolution of comoving black hole accretion rate densities computed according to different prescriptions. The thick solid line shows the results from this work, calculated for $\epsilon = 0.1$ and $x_{cr} = 10^{-3}$. The thin triple dotted-dashed line, instead, shows the results derived from the HXLF alone, assuming radiative efficiency for all sources and a fixed bolometric correction (see e.g. M04). Thin dashed line shows the evolution of the comoving accretion rate density calculated from the evolution of the QSO luminosity function of Boyle et al. (2000) with the bolometric correction of Elvis et al. (1994). Thick dashed and dot–dashed lines in the same plot represent the accretion rate density for sources above and below the critical rate, respectively.

The history of accretion activity is summarized in Fig. 5, where I plot the comoving BHAR density (thick solid line). The thick dashed and dot–dashed lines in the same plot represent the accretion rate density for sources above and below the critical rate, respectively. For comparison, also plotted are the evolution of the comoving accretion rate density calculated by M04 from the HXLF alone, assuming high radiative efficiency for all sources (thin triple dotted-dashed line); and the corresponding quantity, calculated instead from the evolution of the QSO luminosity function of Boyle et al. (2000) with the bolometric correction of Elvis et al. (1994) (thin dashed line). The accretion rate history is therefore dominated by sources with
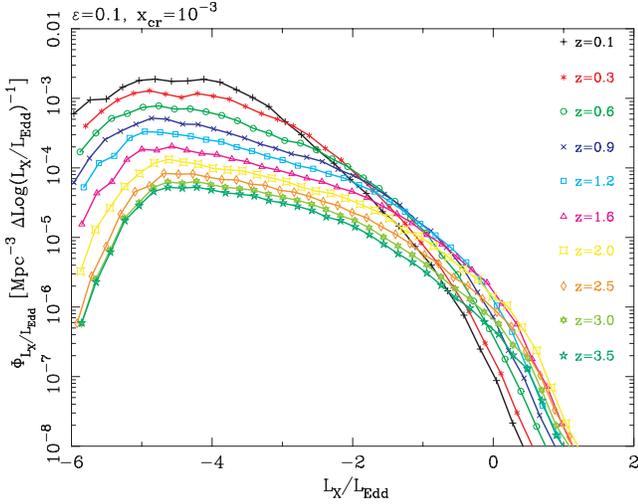
Figure 6. Redshift evolution of the SMBH accretion rate function, from redshift $z = 3.5$ till redshift $z = 0.1$. Different colours and symbols correspond to different redshift bins.

Figure 7. The shaded area is the region of the parameter space (ϵ, x_{cr}) which must be exclude, as the total black hole mass density at $z = 0$ is smaller than the total calculated mass density accreted over cosmic time.

Figure 8. The average radiative efficiency, ϵ_{rad} for the evolving SMBH population between redshift $z = 3.5$ and $z = 0$, as a function of the accretion efficiency ϵ. Shown are the calculations carried out for three values of the critical rate: log (x_{cr}) = −3.5 (dashed line); −3 (fiducial case, solid line) and −2.5 (dot–dashed line). Only values of ϵ that are above the line $\epsilon \simeq 0.2$ $\log(x_{cr}) + 0.7$ (see Fig. 7) are shown.

high radiative efficiency, which explains why many authors were indeed able to explain most of the local SMBH population as remnants of bright AGN phases (see e.g. YT02). However, as a result of the progressive decrease of the average accretion rate with time, as shown in Fig. 3, the population of accreting SMBHs at $z \lesssim 0.5$ is dominated by RIAF sources. For this reason, the evolution of the BHAR density from $z = 0$ to $z = 1$ is less rapid than what would be inferred assuming that all growing SMBH are high-$\dot{m}$ objects. The consequences of this fact for the history of AGN feedback are discussed in Section 6.2.

Finally, in Fig. 6, the evolution of the BHAR (expressed here as the X-ray to Eddington ratio, L_X/L_{Edd}) function is plotted for different redshift intervals, from $z = 0.1$ to $z = 3.5$. While the number of sources accreting at low rates increases monotonically with decreasing redshift, the situation is different for rapidly accreting objects. High accretion rate sources (which should correspond to QSO and bright AGNs) rapidly increase in number with increasing redshift. The cut-off redshift, above which the number of sources declines again, is a function of the typical X-ray to Eddington ratio, being lower for lower accretion rate sources.

The combined evolution of mass and accretion rate functions derived here is caused by the strong trend observed in deep X-ray selected samples (Cowie et al. 2003; Hasinger 2003; Fiore et al. 2003), where progressively lower luminosity AGN reach their maximal space density at progressively lower redshifts.

5.1 Exploring the parameter space: constraints on accretion efficiency

As discussed already, recent works by YT02, Elvis et al. (2002) and M04 have shown that a direct comparison between local SMBH mass density and the AGN/QSO energy density integrated over cosmic time can be used to constrain accretion physics and, in particular, the mean radiative efficiency of all accreting SMBH throughout their history. In doing so, the obvious inequality should be satisfied, that the local $\rho_{BH,0}$ be always larger than (or at most equal to) the total mass density accreted on to active black holes. If this is not the case, the simplest solution to the problem is to assume a higher accretion (and radiative) efficiency, as indeed argued by Elvis et al. (2002) and YT02.
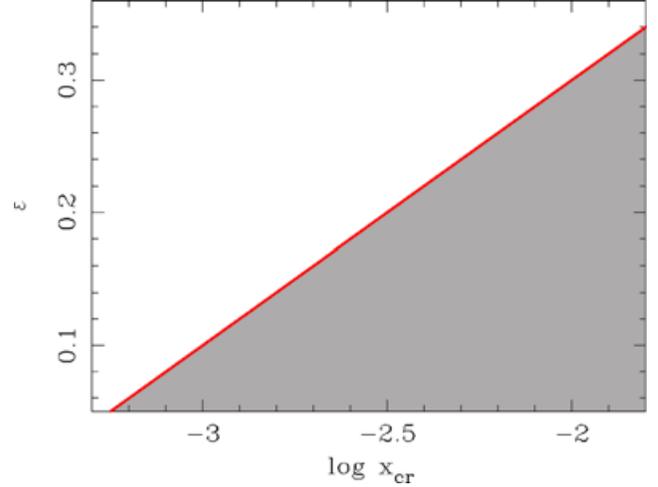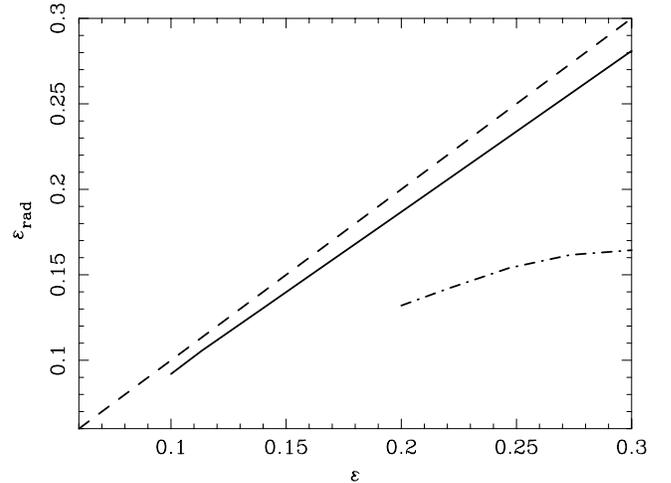
A similar line of reasoning can of course be applied to our calculations. Following the evolution of the BHMF backwards in time, we can search for the values of the free parameters ϵ and x_{cr} for which the SMBH mass density becomes negative in a finite time. For the specific $x(\dot{m})$ and bolometric correction values adopted (see Section 3), I found the acceptable region is bounded below by the following empirical relation: $\epsilon \simeq 0.2 \log(x_{cr}) + 0.7$, which is shown in Fig. 7, with the shaded area representing the excluded region of the parameter space. The mean *radiative* efficiency of SMBH that these calculations yield is itself a function of ϵ and x_{cr}, as is shown in Fig. 8. Obviously, the higher the critical X-ray to Eddington rate where the transition occurs, the lower the average radiative efficiency is for any given value of ϵ, as a larger number of objects of any given L_X will be in the RIAF regime of accretion at any time.

Figs 7 and 8 seem to suggest that a relatively high average accretion efficiency is to be preferred, implying a non-zero average spin

for the SMBH population, unless x_{cr} is lower than 10^{-3}. It should be stressed, however, that such a constraint also depends crucially on the value of the local black hole mass density,[3] and on the adopted bolometric correction, and that the uncertainties in the HXLF and, to a larger extent, in the RLF adopted here should affect the exact determination of the accepted region of the parameter space.

In any case, it is worth stressing that the qualitative behaviour of the BHMF evolution, being driven essentially by the evolution of the shape of the two luminosity functions, is not modified by changes in ϵ and/or x_{cr}. In particular, the anti-hierarchical character of the solution found is a robust result of the approach presented here.

6 DISCUSSION

In Section 5 I have shown and discussed what is possible to deduce about the evolution of the SMBH mass function from the coupled evolution of HXLF and RLF. The limits of this approach are a consequence of the extreme difficulty of obtaining high-redshift luminosity functions in the radio and X-ray bands. Little can be said then, about the evolution of the sources at higher redshifts, where there are hints of a decline of AGN activity with redshift.

Theoretically, the high-z evolution should be mainly driven by the gravitational growth of structures, with the AGN activity triggered by mergers, according to the standard hierarchical picture (Efstathiou & Rees 1988; Cavaliere & Vittorini 2000; Wyithe & Loeb 2003; Di Matteo et al. 2003; Menci et al. 2003). In those phases of early structure formation and evolution, there should be plenty of gas available for accretion owing to the frequent galaxy merging events which can effectively destabilize it and make it reach the central SMBH sphere of influence. Black hole growth is essentially limited by the Eddington limit (the 'feast', according to Small & Blandford 1992, or the 'self-limited regime', Cavaliere & Vittorini 2000). The shape of the BHMF inferred at $z \sim 3$ (see Fig. 4), suggesting that higher mass black holes were already in place at that time, requires a very rapid growth and very high accretion rates in the high-density peaks of the cosmic density distribution at early times, also consistent with the $\dot{m}$ trend shown in Fig. 3.

On the other hand, the rapid, anti-hierarchical, evolution of the SMBH population between $z = 0$ and $z \sim 2.5$ has always been difficult to incorporate into standard CDM models for structure formation. It is interesting, in this context, to compare the main qualitative, anti-hierarchical, picture emerging for growing black holes with the evolution of the galaxy population. As for SMBHs, the most striking and robust evolutionary trend observed for the galaxy population is the rapid increase of the global star formation rate (SFR) between $z = 0$ and $z \sim 1$–2 (Lilly et al. 1996; Madau et al. 1996). Indeed, Di Matteo et al. (2003) and M04 have already shown that the BHAR and SFR histories have broadly similar shapes, with an approximately constant ratio of about a few $\times 10^{-3}$, and this has also been recently measured directly by the SDSS (Heckman et al. 2004). Cowie et al. (2003) showed that such an evolution is caused by the smooth decline with redshift (from $z \sim 1$ to the present) of the rest-frame K band luminosity (an indicator of the total stellar mass) of galaxies that undergo rapid star formation. This phenomenon, called 'down-sizing' is the exact analogous of the anti-hierarchical growth of SMBHs described here, and has been recently confirmed

by Kauffmann et al. (2004), who studied the environmental dependence of the relation between star formation and stellar mass in a large number of SDSS galaxies.

What is the physical origin of such a common behavior? The combined effects of the decrease in the galaxy interaction rate in the era of the formation of groups and clusters (see e.g. Cavaliere & Vittorini 2000), the expansion of the Universe affecting the gas cooling efficiency (Hernquist & Springel 2003; Di Matteo et al. 2003), the progressive depletion of the cold gas reservoirs within galaxies needed to power accretion and strong feedback from both stars and AGN (see e.g. Wyithe & Loeb 2003; Granato et al. 2004), should all contribute to the observed evolution at redshift below 3, which can be defined as that of supply-limited accretion (Cavaliere & Vittorini 2000), or the famine after the feast (Small & Blandford 1992). Clearly, the aim of this paper is not to explore all such theoretical issues and predictions, but rather to provide a detailed picture to compare those predictions with. Therefore, the rest of the discussion will be devoted to two specific results on the lifetimes and duty cycles of active black holes and on the possible implication of the derived SMBH growth history for the AGN feedback evolution.

6.1 Lifetimes and duty cycles of active black holes

In all models that try to derive the properties of the SMBH population from the observed QSO evolution, a key element is represented by the typical quasar lifetime or by the almost equivalent activity duty cycle (see Martini 2004, and references therein). However, the significance of these parameters is limited to the standard case in which, on the basis of an observed luminosity function in a specific waveband, one tries to derive the distribution of either black hole masses or accretion rates. Usually, a constant Eddington ratio is assumed in this case, which implies that QSO are considered as on–off switches. The duty cycle is then simply the fraction of black holes which are active at any time, and the lifetime is the integral of the duty cycle over the age of the Universe.

The picture discussed here is different, in that a broad distribution of Eddington rates is not only allowed, but actually calculated for the SMBH population at every redshift. When this is the case, a more meaningful definition of activity lifetime is needed. I follow the formulation of Steed & Weinberg (2004), by first defining the mean Eddington rate $\langle \dot{m}(M_0, z) \rangle$ for an object of mass M_0 at redshift $z = 0$ and then introducing the mean accretion weighted lifetime of a SMBH with a given mass *today*, i.e.

$$\tau(M_0, z) = \int_{\infty}^{z} \langle \dot{m}(M_0, z') \rangle \frac{dt}{dz'} \, dz'. \tag{12}$$

The ratio of $\tau(M_0, z)$ to the Salpeter time, $t_S = \epsilon Mc^2/L_{Edd} = (\epsilon/0.1) 4.5 \times 10^7$ yr, gives the mean number of e-folds of mass growth for objects with mass M_0 up to redshift z. The ratio of $\tau(M_0, z)$ to the Hubble time $t_{Hubble}(z) = H(z)^{-1}$, instead, is a measure of the activity duty cycle of SMBH.

As I follow here a phenomenological approach based on observed luminosity functions, and no information is therefore available on the formation and early growth of the first black holes, it is interesting here to calculate 'partial' lifetimes in a given redshift interval $\Delta z = (z_i, z_f)$

$$\Delta\tau(M_0, \Delta z) = \int_{z_i}^{z_f} \langle \dot{m}(M_0, z') \rangle \frac{dt}{dz'} \, dz'. \tag{13}$$

In Fig. 9, I show $\Delta\tau(M_0, \Delta z)$ for three redshift intervals: $0 < z \leqslant 1$; $1 < z \leqslant 2$; and $2 < z \leqslant 3$. The accretion weighted lifetime
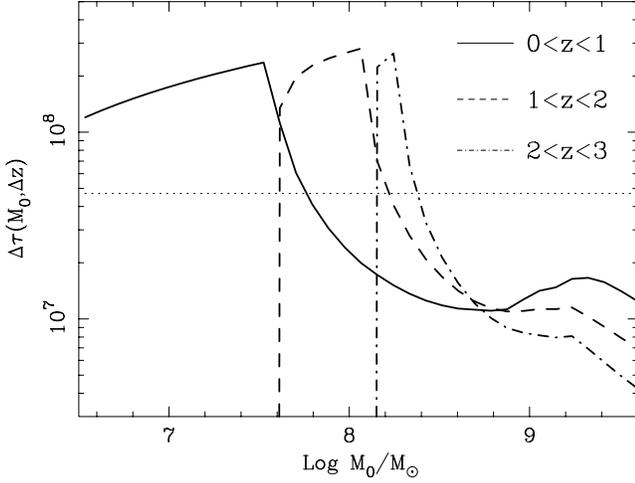

Figure 9. Partial mean accretion weighted lifetimes of SMBH with mass today M_0, calculated for three different redshift intervals: $0 < z < 1$ (solid line), $1 < z < 2$ (dashed line) and $2 < z < 3$ (dot–dashed line). The horizontal dotted line is the Salpeter time for accretion efficiency of 10 per cent. The accretion weighted lifetime for BH of any given mass between $0 < z < 3$ is the sum of the three.

for a black hole of any given mass between $0 < z < 3$ is of course just the sum of the three. The anti-hierarchical nature of mass build-up in actively accreting AGN and QSOs is again clearly illustrated by this plot. In fact, the major growth episode of a SMBH must coincide with the period when $\Delta\tau > t_S$. This happens at $z < 1$ for $M_0 \lesssim 10^{7.6}$, between redshift 1 and 2 for $10^{7.6} \lesssim M_0 \lesssim 10^{8.2}$, and at $2 < z < 3$ for $10^{8.2} \lesssim M_0 \lesssim 10^{8.4}$. SMBHs with masses larger than $M_0 \sim 10^{8.5}$ today, must have experienced their major episodes of growth at redshifts higher than 3. Black holes of lower mass today, which are also accreting at the higher rates in the local universe (see Section 2.1) drop below an hypothetic seed mass (here fixed at $10^4\,M_\odot$, but these results do not depend strongly on this value) and effectively 'disappear' at higher redshifts. This reflects the obvious impossibility of working out the initial condition of black hole growth from the local population evolved backwards once an object has exponentiated its mass just a few times.

It also interesting to note that the objects that dominate the SMBH mass function today, i.e. those in the range of masses around $10^{7.5}\,M_\odot$, where $M_0\,\phi_M(M_0, z = 0)$ peaks, mainly grew around $z \sim 1$, which is when most of the XRB light we see today was emitted (Hasinger 2003).

The ratio $\Delta\tau\,(M_0, \Delta z)/\Delta t(\Delta z)$, where $\Delta t(\Delta z)$ is the time elapsed in the redshift interval Δz, is an indication of the 'partial' duty cycle of a black hole of mass M_0 today in that particular epoch. This is shown in Fig. 10 for the same redshift intervals of Fig. 9.

6.2 On the kinetic energy output and the history of AGN feedback

The main novelty of the constraints on black hole growth history presented here is the opportunity to determine both mass and accretion rate for each and every source, thanks to the fundamental plane relationship. As I discussed in Section 3, different modes of accretion are expected at different $\dot{m}$ values. In fact, it is not just the radiative output of an accreting black hole that scales differently with $\dot{m}$ for different accretion modes, but also the total kinetic energy carried by the jets/outflows that are responsible for radio emission.
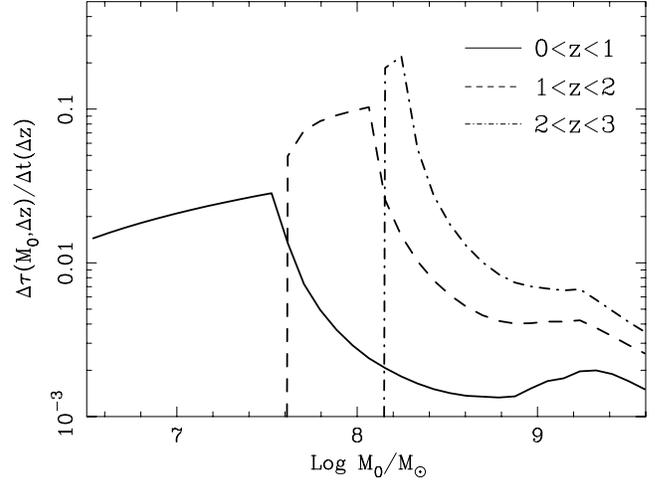

Figure 10. Partial mean accretion weighted duty cycles of SMBH with mass today M_0, calculated for three different redshift intervals: $0 < z < 1$ (solid line); $1 < z < 2$ (dashed line); and $2 < z < 3$ (dot–dashed line).

It is possible, therefore, to derive a parallel history of the (mostly unseen) mechanical power output from the growth of SMBHs.

To derive a scaling of the jet kinetic power, $W_{\rm jet}$, with mass and accretion rate for sources above and below $x_{\rm cr}$, I will proceed in the following way. I will fist assume that the jet kinetic power at injection is carried by internal energy, and I will assume equipartition between the jet magnetic field and the total pressure in the disc (see Heinz & Sunyaev 2003). Then $W_{\rm jet} \propto P_{\rm rel}R_S^2 \propto B^2 M^2$, where $P_{\rm rel}$ represents the pressure in relativistic particles at the base of the jet and $R_S = 2GM/c^2$ is the Schwarzschild radius. The scaling of the magnetic field, instead, can be directly inferred from the inclination of the fundamental plane. Equation (11) of MHD03 relates the observed correlation coefficients of the fundamental plane relation to the slope of the electron distribution in the jet, p, the observable radio spectral index α_R, the logarithmic derivatives of the magnetic field intensity with respect to mass and accretion rate and the index q of the $L_X - \dot{m}$ relation (see Section 3). From that equation we have, assuming $p = 2$ and $\alpha_R = 0$

$$\frac{\partial \ln B}{\partial \dot{m}} = 0.35\,q\,\xi_{RX} \simeq 0.21\,q$$

$$\frac{\partial \ln B}{\partial M} = 0.35(\xi_{RX} + \xi_{RM}) - 1 \simeq -0.51, \qquad (14)$$

where we have used the results of MHD03 $\xi_{RX} = 0.6$ and $\xi_{RM} = 0.78$.

Thus, we obtain the expected result that for RIAF flows ($q \simeq 2.3$) the total jet kinetic power is proportional to only the physical accretion rate $W_{\rm jet} \propto \dot{m}M \propto \dot{M}$ (see Falcke & Biermann 1996; Heinz & Sunyaev 2003; Fender et al. 2003). Whether such an output dominates the energy budget (according to the so-called ADIOS picture, Blandford & Begelman (1999), and as proposed also by Fender et al. 2003) or not depends, however, on the dynamics of the innermost disc–jet coupling and, from the observationally point of view, on the radiative efficiency of the jets. On the other hand, SMBH accreting above the critical rate $x_{\rm cr}$, for which we have assumed $q = 0.76$, should obey the scaling $W_{\rm jet} \propto M\dot{m}^{0.33}$.

Given the above scaling relations for $W_{\rm jet}$, it is possible to calculate the mechanical power output from each accreting black hole and the total integrated one, as a function of redshift. For the sake of simplicity I assume that indeed the total kinetic power of the jet/outflow from RIAF black holes dominates as a sink of energy.
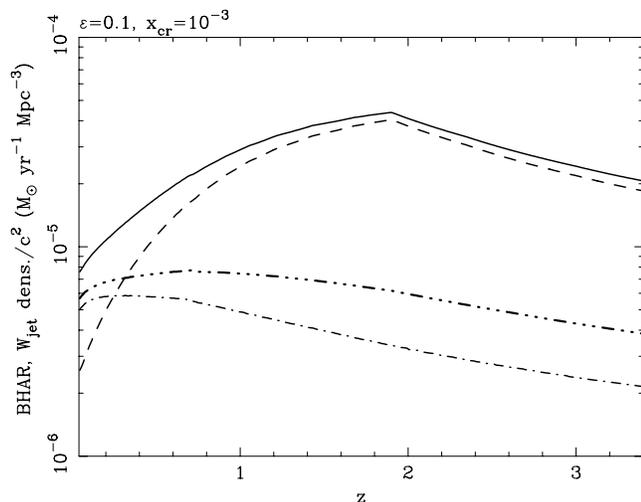
Figure 11. Redshift evolution of the total integrated mechanical power W_{jet}/c^2 from accreting black holes per unit comoving Mpc (triple dotted-dashed line). For comparison, the evolution of the total BHAR density (solid line) is shown together with the two separate contributions from the sources accreting above (dashed line) and below (dot–dashed line) the critical rate, all taken from Fig. 5.

Therefore, the calculated W_{jet} should be considered an absolute upper limit to the jet/outflow kinetic power. The results are shown in Fig. 11.

As the relative contribution of the mechanical energy output is much larger in low accretion rate sources, the history of W_{jet} does not follow that of most luminous sources, but instead exhibits a much weaker evolution, at both low and high redshifts. In addition, because the number of low accretion rate sources increase with time (because of the overall decrease of the average accretion rate, see Section 5), the peak of the mechanical power output lies at a much lower redshift than the peak of the radiative energy output history of AGN, which should have interesting implications for the formation and dynamical evolution of clusters of galaxies.

7 CONCLUSIONS

I have presented a new method to study the growth of accreting SMBHs, based on the simultaneous evolution of the AGN radio and hard (2–10 keV) X-ray luminosity functions. The method is based on the locally observed trivariate correlation between black hole mass, X-ray and radio luminosity (the so-called fundamental plane of black hole activity, MHD03). Thanks to this correlation, it is possible for the first time to break the degeneracy between luminosity, mass and accretion rate that has affected all previous attempts to study the evolution of the SMBHs population by looking at the evolution of a single AGN luminosity function.

Here, the redshift evolution of the SMBH mass function between $z = 0$ and $z \sim 3$ is evaluated integrating backwards in time a continuity equation for the black hole population in which the role of mergers is neglected. The local BHMF, estimated by applying the correlation between black hole mass and velocity dispersion to the velocity distribution function of local galaxies, is used as a boundary condition for the continuity equation. The solution to this equation is uniquely determined once the accretion efficiency is specified, together with the function $L_X/L_{Edd}(\dot{m})$, that links the observed Eddington-scaled hard X-ray luminosity of an accreting black hole to its accretion rate. For the latter, I have assumed a very general form of a double power law, corresponding to a radiatively

inefficient regime at low accretion rates, and a radiatively efficient one at high-$\dot{m}$ values.

By comparing the local observed BHMF with the total mass accreted on to SMBH over their history, it is possible to put simultaneous constraints on the accretion efficiency and on the critical value of the accretion rate, x_{cr}, at which the transition takes place between the two accretion modes.

The main results of this work are the following. For fiducial values of the parameters ($\epsilon = 0.1$ and $x_{cr} = 10^{-3}$), half (~85 per cent) of the local black hole mass density was accumulated at redshift $z < 1$ ($z < 3$), mostly in radiatively efficient episodes of accretion. Qualitatively (i.e. independently of the values of these two parameters), the evolution of the BHMF between $z = 0$ and $z \sim 3$ shows clear signs of anti-hierarchical behaviour. This is a purely phenomenological assessment, and reflects the fact that, while the majority of the most massive objects ($M \gtrsim 10^9$) were already in place at $z \sim 3$, lower mass ones mainly grew at progressively lower redshift, so that the average black hole mass increases with increasing redshift. On the other hand, the average accretion rate decreases towards lower redshift. Therefore, sources in the RIAF regime of accretion only begin to dominate the comoving accretion energy density in the Universe at $z < 1$ (with the exact value of z depending on x_{cr}), while at the peak of the BHAR history, radiatively efficient accretion dominates by almost an order of magnitude. By carefully evaluating the contributions to the total black hole accretion energy density from the different modes of accretion as a function of redshift, I show how to derive a physically motivated AGN feedback history. This, as well as the anti-hierarchical behaviour of SMBH growth described already, may be of some importance for cosmological models of structure formation in the Universe.

REFERENCES

Alexander D. M., Brandt W. N., Hornschemeier A. E., Garmire G. P., Schneider D. P., Bauer F. E., 2001, AJ, 122, 2156
Aller M. C., Richstone D., 2002, AJ, 124, 3035
Barger A. J., Cowie L. L., Brandt W. N., Capak P., Garmire G. P., Hornschemeier A. E., Steffen A. T., Wehner E. H., 2002, AJ, 121, 662
Barger A. J. et al., 2003, AJ, 126, 632
Blandford R. D., Begelman M. C., 1999, MNRAS, 303, L1
Boyle B. J., Shanks T., Croom S. M., Smith R. J., Miller L., Loaring N., Heymans C., 2000, MNRAS, 317, 1014
Bromley J. M., Somerville R. S., Fabian A. C., 2004, MNRAS, 350, 456
Cattaneo A., Bernardi M., 2003, MNRAS, 344, 45
Cattaneo A., Haehnelt M., Rees M. J., 1999, MNRAS, 308, 77
Cavaliere A., Vittorini V., 2000, ApJ, 543, 599
Cavaliere A., Vittorini V., 2002, ApJ, 570, 114
Cowie L. L., Barger A. J., Bautz M. W., Brandt W. N., Garmire G. P., 2003, ApJ, 584, L57
Di Matteo T., Croft R. A. C., Springel V., Hernquist L., 2003, ApJ, 593, 56
Dunlop J. S., 1998, in Bremer M. N., Jackson N., Perez-Fournon I., eds, Observational Cosmology with the New Radio Surveys. Kluwer, Dordrecht, p. 157
Dunlop J. S., Peacock J. A., 1990, MNRAS, 247, 19
Efstathiou G., Rees M. J., 1988, MNRAS, 230, 5
Elvis M. et al., 1994, ApJS, 95, 1
Elvis M., Risaliti G., Zamorani G., 2002, ApJ, 565, L75
Fabian A. C., 1999, MNRAS, 308, L39

Fabian A. C., 2004, in Ho L., ed., Carnegie Observatories Astrophysics Series Vol. 1, Coevolution of Black Holes and Galaxies. Cambridge Univ. Press, Cambridge, p. 447

Fabian A. C., Iwasawa K., 1999, MNRAS, 303, L34

Falcke H., Biermann P. L., 1996, A&A, 308, 321

Fan X. et al., 2001, AJ, 121, 54

Fender R. P., Gallo E., Jonker P. G., 2003, MNRAS, 343, L99

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Fiore F. et al., 2003, A&A, 409, 79

Gandhi P., Fabian A. C., 2003, 339, 1095

Gebhardt K. et al., 2000, ApJ, 539, L13

Granato G. L., Silva L., Monaco P., Panuzzo P., Salucci P., De Zotti G., Danese L., 2001, MNRAS, 324, 757

Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580

Haehnelt M., Rees M. J., 1993, MNRAS, 263, 168

Haehnelt M., Natarajan P., Rees M. J., 1998, MNRAS, 300, 817

Haiman Z., Loeb A., 1998, ApJ, 503, 505

Haiman Z., Menou K., 2002, ApJ, 531, 42

Hasinger G., 2003, AIP Conf. Proc. 666, 227

Hatziminaoglou E., Mathez G., Solanes J.-M., Manrique A., Salvador-Solé , 2003, MNRAS, 343, 692

Heckman T. M., Kauffmann G., Brinchmann J., Charlot S., Tremonti C., White S., 2004, ApJ, in press (astro-ph/0406218)

Heinz S., Sunyaev R. A., 2003, MNRAS, 343, L59

Hernquist L., Springel V., 2003, MNRAS, 34, 1253

Hosokawa T., 2004, ApJ, 606, 139

Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, in press (doi:10.1111/j.1365-2966.2004.08117.x(astro-ph/0402030)

Kormendy J., Richstone D., 1995, ARA&A, 33, 581

Laor A., 1990, MNRAS, 246, 369

Lilly S. J., Le Fevre O., Hammer F., Crampton D., 1996, ApJ, 460, L1

Maccarone T., 2003, A&A, 409, 697

Maccarone T., Gallo E., Fender R. P., 2003, MNRAS, 345, L19

Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388

Magorrian J. et al., 1998, AJ, 115, 2285

Mahmood A., Devriendt J. E., Silk J., 2004, preprint (astro-ph/0401003)

Mainieri V., Bergeron J., Hasinger G., Lehmann I., Rosati P., Schmidt P., Szokoly G., Della Ceca R., 2002, A&A, 338, 781

Malkan M. A., 1983, ApJ, 268, 582

Marconi A., Hunt L. K., 2003, ApJ, 589, L21

Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169 (M04)

Martini P., 2004, in Ho L. ed., Carnegie Observatories Astrophysics Series Vol. 1, Coevolution of Black Holes and Galaxies. Cambridge Univ. Press, Cambridge, p. 170.

Marzke R. O., Geller M. J., Uchra J. P., Corwin H. G., 1994, AJ, 108, 437

McClintock J. E., Remillard R. A., 2004, in Lewin W. H. G., van der Klis M., eds, Compact Stellar X-ray Sources. CAmbridge Univ. Press, Cambridge, in press (astro-ph/0306213)

McLure M. J., Dunlop J. S., 2002, MNRAS, 331, 795

McLure M. J., Dunlop J. S., 2004, MNRAS, in press (doi:10.1111/j.1365-2966.2004.08034.x)(astro-ph/0310267)

Menci N., Cavaliere A., Fontana A., Giallongo E., Poli F., Vittorini V., 2003, ApJ, 587, L63

Menou K., Haiman Z., 2004, ApJ, in press (astro-ph/0405335)

Merloni A., 2003, MNRAS, 341, 1051

Merloni A., Heinz S., Di Matteo T., 2003, MNRAS, 345, 1057 (MHD03)

Merritt D., Ferrarese L., 2001, MNRAS, 320, L30

Monaco P., Salucci P., Danese L., 2000, MNRAS, 311, 279

Narayan R., 2002, in Gilfanov M., Sunyaev R., Churazov E., eds, Proceedings of Lighthouses of the Universe. Springer-Verlag, Berlin, p. 405

Novikov I. D., Thorne K. S., 1973, in De Witt C., De Witt B., eds, Black Holes. Gordon & Breach, New York, p. 343

Page M. J., 2001, MNRAS, 328, 925

Quataert E., 2003, Astron. Nachr., 324, 435

Rees M. J., Phinney E. S., Begelman M. C., Blandford R. D., 1982, Nature, 295, 17

Richstone D. et al., 1998, Nat, 395, 14

Rosati P. et al., 2002, ApJ, 566, 667

Sadler E. M. et al., 2002, MNRAS, 329, 227

Salucci P., Szuszkiewicz E., Monaco P., Danese L., 1999, MNRAS, 307, 637

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Silk J., Rees M. J., 1998, A&A, 331, L1

Small T. A., Blandford R. D., 1992, MNRAS, 259, 725

Soltan A., 1982, MNRAS, 200, 115

Spergel D. N. et al., 2003, ApJS, 148, 175

Steed A., Weinberg D. H., 2004, ApJ, submitted (astro-ph/0311312)

Tremaine S. et al., 2002, ApJ, 574, 740

Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886

Umemura M., 2001, ApJ, 560, L29

Vestergaard M., 2004, ApJ, 601, 676

Vignali C., Brandt W. N., Schneider D. P., 2003, AJ, 125, 433

Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559

Willott C. J., Rawlings S., Blundell K. M., Lacy M., Eales S. A., 2001, MNRAS, 322, 536

Wolf C., Wisotzki L., Borch A., Dye S., Kleinheinrich M., Meisenheimer K., 2003, A&A, 408, 499

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

Yu Q., Lu Y., 2004, ApJ, 602, 603

Yu Q., Tremaine S., 2002, MNRAS, 335, 965 (YT02)

APPENDIX A: ANALYTICAL APPROXIMATION OF THE X-RAY LUMINOSITY FUNCTION

I adopt here the functional form for the HXLF described in Ueda et al. (2003), which has the following analytic approximation:

$$\phi_X(L_X, z) = \phi_X(L_X, 0)\eta(z, L_X), \tag{A1}$$

where the local X-ray luminosity function is expressed as a smoothly connected double power law

$$\phi_X(L_X, z = 0) = A\left[(L_X/L_*)^{\gamma_1} + (L_X/L_*)^{\gamma_2}\right]^{-1}, \tag{A2}$$

while the evolutionary part is expressed as

$$\eta(z, L_x) = \begin{cases} (1 + z)^{p_1} & z < z_c(L_X) \\ \eta(z_c)[(1 + z)/(1 + z_c(L_X))]^{p_2} & z \geqslant z_c(L_X) \end{cases} \tag{A3}$$

and

$$z_c(L_x) = \begin{cases} z_c^* & L_X \geqslant L_a \\ z_c^*(L_X/L_a)^\alpha & L_X < L_a. \end{cases} \tag{A4}$$

For a ΛCDM universe, the best-fitting parameters are: $A = 5.04 \pm 0.33 \times 10^{-6} h_{70}^3$ Mpc^{-3}; $L_* = 10^{(43.94^{+0.21}_{-0.26})} h_{70}^{-2}$ erg s^{-1}; $\gamma_1 = 0.86 \pm 0.15$; $\gamma_2 = 2.23 \pm 0.13$; $p_1 = 4.23 \pm 0.39$; $p_2 = -1.5$ (fixed); $z*_c = 1.9$ (fixed); $L_a = 10^{44.6} h_{70}^{-2}$ erg s^{-1} (fixed); $\alpha = 0.335 \pm 0.070$.

APPENDIX B: ANALYTICAL APPROXIMATION OF THE RADIO LUMINOSITY FUNCTION

The 5-GHz RLF is derived from the low-frequency (151-MHz) function of Willott et al. (2001), assuming a fixed radio spectral index of $\alpha_R = 0.7$ to rescale the luminosities.

The analytical approximation is that of model C of Willott et al. (2001), and is given by a sum of two differently evolving populations (a low- and a high-luminosity one), i.e.

$$\phi_{R,151}(L_{151}, z) = \phi_{R,l} + \phi_{R,h}. \tag{B1}$$

These two populations evolve differently with redshift, according to the following expressions:

$$\phi_{R,l} = \begin{cases} \phi^0_{R,l} \left(\frac{L_{151}}{L_{l,*}} \right)^{-\alpha_l} \exp\left(\frac{-L_{151}}{L_{l,*}} \right) (1+z)^{k_l}; & z < z_{l0} \\ \phi^0_{R,l} \left(\frac{L_{151}}{L_{l,*}} \right)^{-\alpha_l} \exp\left(\frac{-L_{151}}{L_{l,*}} \right) (1+z_{l0})^{k_l}; & z \geqslant z_{l0} \end{cases} \tag{B2}$$

$$\phi_{R,h} = \phi^0_{R,h} \left(\frac{L_{151}}{L_{h,*}} \right)^{-\alpha_h} \exp\left(\frac{-L_{h,*}}{L_{151}} \right) f_h(z). \tag{B3}$$

The high-luminosity redshift evolution $f_h(z)$ is given by

$$f_h(z) = \begin{cases} \exp\left[-\frac{1}{2} \left(\frac{z - z_{h0}}{z_{h1}} \right)^2 \right] & z < z_{h0} \\ \exp\left[-\frac{1}{2} \left(\frac{z - z_{h0}}{z_{h2}} \right)^2 \right] & z \geqslant z_{h0}. \end{cases} \tag{B4}$$

The best-fitting parameters to the observed luminosity function are: $\log \phi^0_{R,l} = -7.12^{+0.10}_{-0.11}, \alpha_l = 0.539 \pm 0.02, \log L_{l,*} = 26.10^{+0.08}_{-0.09}, z_{l0} = 0.71 \pm 0.10, k_l = 4.30^{+0.57}_{-0.55}, \log \phi^0_{R,h} = -6.20^{+0.09}_{-0.11}, \alpha_h = 2.27^{+0.12}_{-0.11}, \log L_{h,*} = 26.95^{+0.11}_{-0.10}, z_{h0} = 1.91 \pm 0.16, z_{h1} = 0.56 \pm 0.05, z_{h2} = 1.38^{+0.52}_{0.28}$.

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