Distant monsters: high redshift AGN predictions from a ΛCDM cosmological model

Andrew J. Griffin,1⋆ Cedric G. Lacey,1 Violeta Gonzalez-Perez,1,2 Claudia del P. Lagos,3,4 Carlton M. Baugh1 Nikos Fanidakis,5,6

1 Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK
2 Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth PO1 3FX, UK
3 International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
4 ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)
5 Max-Planck-Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
6 BASF, Carl-Bosch Strasse 38, 67056 Ludwigshafen, Germany

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The next decade promises to revolutionise the study of the high redshift Universe: upcoming missions such as JWST, EUCLID and ATHENA will greatly advance our understanding of the early stages of galaxy formation. We use an updated model for the evolution of masses and spins of supermassive black holes (SMBHs), coupled to the latest version of the semi-analytical model of galaxy formation galform using the Planck cosmology, and a new high resolution Millennium style dark matter simulation to make predictions for multiwavelength Active Galactic Nuclei (AGN) luminosity functions in the redshift range 0 < z < 15. We compare the model to the observed black hole mass function and the SMBH versus galaxy bulge stellar mass relation at z = 0, and to the predicted bolometric, hard X-ray, soft X-ray and optical AGN luminosity functions to observations at z < 6, and find that the model is in good agreement with the observations. We employ this successful model to make predictions for the AGN luminosity function at z > 6 in the broadband filters of JWST, EUCLID and ATHENA. We predict that these three surveys will select three somewhat different samples of SMBHs, with EUCLID unveiling the more massive high accretion rate SMBHs, JWST the less massive, lower accretion rate SMBHs and ATHENA covering objects inbetween. Overall, typical detected SMBHs will have masses, $M_{\text{BH}} \sim 10^6 - 10^7 M_\odot$, with dimensionless mass accretion rates, $\dot{M}/\dot{M}_{\text{Edd}} \sim 1 - 10$, in host galaxies of stellar mass, $M_\star \sim 10^8 - 10^9 M_\odot$. The future comparison of these predictions to observations will provide insights into our understanding of SMBHs in galaxy formation.

Key words: galaxies: high-redshift – galaxies: active – quasars: general

1 INTRODUCTION

Ever since quasars were first identified to be cosmological sources (Schmidt 1968), a key aim has been to to understand their evolution through cosmological time. Early studies showed that the number density of quasars shows strong evolution, with more luminous quasars present at z ≈ 2 than at z ≈ 0, leading to the suggestion that quasars evolve by ‘pure luminosity evolution’ (PLE). In this scenario, quasars are long lived and fade through cosmic time, leading to an evolution in the luminosity function of luminosity only (e.g. Boyle et al. 1990). However, more recent optical surveys, which probe brighter objects, have shown not only that the slope of the luminosity function evolves (e.g. Richards et al. 2006; Croom et al. 2009), but also that the number density decreases at high redshift (e.g. Fan et al. 2001; Ji-ang et al. 2016). Surveys at X-ray wavelengths, which detect lower luminosity objects, also show a more complicated evolution than PLE, showing an evolution in the shape of the luminosity function (e.g. Ueda et al. 2014) and differences between the absorbed and unabsorbed populations (e.g. Aird et al. 2015). Clearly, the full picture of supermassive black holes (SMBHs) and Active Galactic Nuclei (AGN) evolving through cosmological time is complicated, and requires detailed investigation. Cosmological simula-

⋆ E-mail: andrew.j.griffin@durham.ac.uk (AJG)
tions have allowed us to try to quantify the role of different contributing accretion mechanisms (e.g. mergers, disc instabilities) and obscuration to the AGN luminosity function (e.g. Fanidakis et al. 2012; Hirschmann et al. 2012), but we do not fully understand the reasons for the different features of the evolution.

The evolution of AGN through time also has significant impact on galaxy evolution, since AGN are inferred to have a dramatic effect on their host galaxies. The relativistic jets from AGN can have a strong effect on the host galaxy by driving out gas to form huge X-ray cavities (e.g. Forman et al. 2005; Randall et al. 2011; Blanton et al. 2011), or the AGN can drive powerful high-velocity outflows (e.g. Pounds et al. 2003; Reeves et al. 2003; Rupke & Veilleux 2011). Powerful AGN are also observed at higher redshift, with large scale outflows driven by AGN observed at $z \sim 2$ (Harrison et al. 2012), and at $z \sim 6$ (Maiolino et al. 2012; Cicone et al. 2015). Recently, Bañados et al. (2018) reported observations of the most distant quasar to date at $z = 7.64$, and the multiwavelength analysis of these high redshift QSOs have revealed puzzling results, such as the high incidence of highly starbursting companions (Decarli et al. 2017). In addition, X-ray observations have shown that faint QSOs may play an important role in reionising the Universe (Giallongo et al. 2015).

The next decade promises to be an exciting one with regard to understanding the high redshift Universe. The launch of James Webb Space Telescope (JWST) in 2020 will pave the way for an increased understanding of the $z > 7$ Universe (e.g. Gardner et al. 2006). JWST will make observations from the optical to near-infrared (0.7 $\mu$m to 30 $\mu$m) to probe the earliest galaxies and the stars contained within. The science objectives of the Advanced Telescope for High-ENERGY Astrophysics (ATHENA) are to determine SMBH seeds and investigate the influence of SMBHs on the formation of the first galaxies (Nandra et al. 2013), by making observations at X-ray wavelengths (0.5-10 keV). While EUCLID is primarily a cosmology mission with the aim of constraining dark energy, its optical to near-infrared (0.5-2 $\mu$m) coverage means it will also detect high-redshift quasars at $z > 8$ (Laurejs et al. 2011). We are now entering an era where the role of SMBHs in the high redshift Universe can be robustly probed.

The precise physical mechanism of the production of AGN jets has not yet been determined, but the two most popular mechanisms are either that the accretion flow determines the jet power (Blandford & Payne 1982) or that the spin of the black hole determines the jet power (Blandford & Znajek 1977). Many simulations have been conducted to study jet formation, where black hole spin often plays a key role (e.g. Kudoh et al. 1998; Hawley & Balbus 2002; McKinney 2005; Hawley & Krolik 2006; Tchekhovskoy & McKinney 2012; Sadowski et al. 2013). Given that black hole spin is closely related to the production of AGN jets, observational studies have attempted to measure black hole spin values (Brenneman & Reynolds 2006; Chiang & Fabian 2011; Done et al. 2013) and perform cosmological simulations of black hole spin evolution (Berti & Volonteri 2008; Lagos et al. 2009; Fanidakis et al. 2011; Barausse 2012; Dotti et al. 2013). The role of the latter has been to understand the role of black hole mergers and accretion across cosmological time on SMBH growth and spin evolution.

Constructing a model that accounts for all the processes involved in changing SMBH spin is not a simple task - especially given the vast range of scales involved. On sub-parsec scales, the spin of the black hole affects the radius of the last stable orbit for orbiting material, and hence the radiative efficiency of the black hole. The SMBH can be misaligned with the accretion disc, causing the accretion disc to become warped which affects how gas is accreted. On larger (parsec to kiloparsec) scales it is currently unclear whether the gas accretes in a coherent manner or in a series of randomly oriented episodes (King et al. 2008), which affects how much the black hole spins up. On kiloparsec scales, the SMBH can accrete gas from cold, infalling gas (made available from galaxy mergers or disc instabilities) or from gas accreted from the hot halo gas surrounding the galaxy. The SMBH may also drive a kiloparsec scale jet - strongly affecting the behaviour of the host galaxy, and hence the gas that infalls onto the SMBH.

To be able to model all these processes on all these scales, some form of 'sub-grid' prescription is required to be able to model the effects going on below the numerical resolution of the calculation. Therefore, investigating SMBH spin evolution and SMBH mass growth and exploring its effect on galaxy wide scales is well suited to using a semi-analytic model of galaxy formation. Using a semi-analytic model of galaxy formation coupled with a high-volume, high-resolution dark matter simulation means that we can conduct detailed simulations within a computationally reasonable time-frame, which means that not only can we investigate the low redshift Universe with greater accuracy, but also we can make more accurate predictions for the high redshift Universe than ever before.

Semi-analytic models of galaxy formation have greatly contributed to our understanding of SMBHs and AGN in galaxy formation. Bower et al. (2006) used galform (one such model) in which a relativistic jet balances radiative cooling in the most massive haloes provides a match to the galaxy luminosity function at a range of redshifts, highlighting the potential importance of AGN feedback in galaxy formation. Malbon et al. (2007) extended the galform model of Baugh et al. (2005) by including SMBH growth from mergers, cold gas accreted from starbursts and from the hot halo mode introduced in Bower et al. (2006) to reproduce the quasar optical luminosity function. In Fanidakis et al. (2011), galform was updated to include an SMBH spin evolution model in which SMBH spin evolves during accretion of gas or by merging with other SMBHs. This model was then compared to observed AGN luminosity functions in the redshift range $0 < z < 6$ for optical and X-ray data in Fanidakis et al. (2012). Other semi-analytic galaxy formation models have also investigated SMBH growth and evolution (e.g. Lagos et al. 2008; Marulli et al. 2008; Bonoli et al. 2009; Hirschmann et al. 2012; Menci et al. 2013; Neistein & Netzer 2014; Enoki et al. 2014; Shirakata et al. 2018) and studies have also been conducted using hydrodynamical simulations (e.g. Hirschmann et al. 2014; Sijacki et al. 2015; Rosas-Guevara et al. 2016; Volonteri et al. 2016).

In this paper, we present predictions for higher redshifts than these previous studies, as we make predictions for $z > 6$, for comparison with observations from JWST, EUCLID and ATHENA. We also include a more detailed treatment of
the obscuration and compare the model predictions to more recent observational data.

This paper is organised as follows. In Section 2 we outline the model used. In Section 3 we show some basic black hole plots (the black hole mass function, black hole mass versus bulge/galaxy stellar mass, black hole spin distributions) for the model. In Section 4 we show the evolution of the AGN luminosity function at different wavelengths for $0 < z < 6$. In Section 5 we present predictions for the AGN luminosity function for up to $z = 15$ and in Section 6 we give concluding remarks.

2 THE MODEL

To construct our predictions, we use the Durham semi-analytic model of galaxy formation, GALFORM. Building on the principles outlined in White & Rees (1978), White & Frenk (1991), Cole et al. (1994) and introduced in Cole et al. (2000), in GALFORM galaxies form from baryons condensing within dark matter haloes, with the assembly of the haloes described by the dark matter halo merger trees. While dark matter merger trees can be calculated using a Monte-Carlo technique that is based on the Extended Press-Schechter theory (Lacey & Cole 1993; Cole et al. 2000; Parkinson et al. 2008) they can also be extracted from dark matter N-body simulations (Kauffmann et al. 1999; Helly et al. 2003; Jing et al. 2014), which is the method that we follow in this paper. The baryonic physics is then modelled using a set of coupled differential equations to track the exchange of baryons between different galaxy components. The physical processes modelled in GALFORM include: i) the merging of dark matter haloes, ii) shock heating and radiative cooling of gas in haloes, iii) star formation from cold gas, iv) photoionisation/supernova/AGN feedback, v) the chemical evolution of gas and stars, vi) galaxies merging in haloes due to dynamical friction, vii) the evolution of stellar populations, and viii) the extinction and reprocessing of stellar radiation by dust. For a detailed description of the physical processes involved, see Lacey et al. (2016) and references therein.

In this paper we update the model for SMBHs and AGN presented in Fanidakis et al. (2011), superceding the equations in that paper, which contained some typographical errors, and also putting special emphasis on improving the model for the obscuration of AGN at X-ray and optical AGN wavelengths. We incorporate the updated Fanidakis et al. (2011) SMBH model in the Lacey et al. (2016) GALFORM model. The Lacey et al. (2016) model brings together several GALFORM developments into a single model, which fits well a wide range of observational data covering wavelengths from the far-UV to the sub-mm in the redshift range $0 < z < 6$. The Lacey et al. (2016) GALFORM model differs in a number of ways from that used in Fanidakis et al. (2011, 2012), including having different IMFs for quiescent and starburst star formation, as opposed to the single IMF used in Fanidakis et al. (2011). In this analysis, we extend the model predictions beyond the $z = 6$ limit in Fanidakis et al. (2012).

The dark matter simulation used for these predictions is a new (800 Mpc)$^3$ Millennium style simulation (Springel et al. 2005) with cosmological parameters consistent with the Planck satellite results (Planck Collaboration et al. 2016) - henceforth referred to as the P-Millennium (Baugh et al. in prep.). The P-Millennium has an increased number of snapshots output - 270 instead of 64 for the Millennium simulation, and a halo mass resolution of $2.12 \times 10^9 M_{\odot}$ compared to the halo mass resolution of $1.87 \times 10^{10} M_{\odot}$ for the dark matter simulation used in Lacey et al. (2016). Given that in the ‘Dhalo’ halo finding algorithm used (Jing et al. 2014), a halo is required to have 20 dark matter particles, P-Millennium has a dark matter particle mass of $1.06 \times 10^9 M_{\odot}$.

Because of the changed cosmological parameters and improved halo mass resolution in P-Millennium compared to the simulation used in Lacey et al. (2016), it is necessary to re-calibrate some of the galaxy formation parameters - this is discussed more thoroughly in Baugh et al. (in prep.). The parameters that have been changed from the Lacey et al. (2016) model values are shown in Table 1. The model also includes a more accurate calculation of the merging of satellite galaxies (Simha & Cole 2017). This P-Millennium based model has already been used in Cowley et al. (2018) to make predictions for galaxies for JWST in near- and mid-IR bands, and a model using P-Millennium and the model of Gonzalez-Perez et al. (2018) was used to study the effect of AGN feedback on halo occupation distribution models in McCullagh et al. (2017).

SMBHs in GALFORM grow in three different ways. Firstly, SMBHs can accrete gas during starbursts, which are triggered by either galaxy mergers or disc instabilities. The galaxy merger causes gas to be transferred to the centre and triggers a burst of star formation (e.g. Mihos & Hernquist 1996). Some of this gas is then available to feed the central SMBH (Kauffmann & Haehnelt 2000; Malbon et al. 2007). Disc instabilities cause a bar to be formed, which disrupts the galaxy disc (Elstathen et al. 1982) which transfers gas to the centre to be fed into the SMBH. Secondly, SMBH mass can be built up by SMBH-SMBH mergers. This is representative of the processes that occur when galaxies merge: dynamical friction from gas and stars causes the SMBH of the smaller galaxy to sink towards the other SMBH. Then, as the separation decreases, gravitational radiation provides a mechanism by which the SMBHs can lose angular momentum and spiral in to merge and form a larger SMBH. Thirdly, SMBHs can accrete gas from the hot gas atmospheres of massive haloes: when large haloes collapse, gas is shock heated to form a quasistatic halo atmosphere. For sufficiently massive haloes, the cooling time of this gas is longer than its free-fall time and so the SMBH is fed with a slow inflow from the halo’s hot atmosphere - ‘hot halo mode accretion’ (Bower et al. 2006).

2.1 SMBH seeds

The starting point for the treatment of SMBHs in the model is as SMBH seeds that will eventually grow by accretion of gas and by merging with other SMBHs to form the objects in the Universe today. The processes for SMBH seed formation are uncertain (see e.g. Volonteri 2010, and references therein) and so we simply add a seed SMBH of mass $M_{\text{seed}}$ into each halo, where $M_{\text{seed}}$ is a parameter that we can vary. Unless otherwise stated, this parameter has the value $M_{\text{seed}} = 10 h^{-1} M_{\odot}$ - representative of the SMBH seed.
formed by stellar collapse. The effect of varying this seed mass is discussed in Appendix A.

### 2.2 SMBH mass growth and spinup by gas accretion

The model used takes into account the SMBH spin. In this model, SMBHs can change spin in two different ways, by accretion of gas or by merging with another SMBH. The SMBH spin is characterised by the dimensionless spin parameter, \( a = cJ_{\text{BH}}/GM_{\text{BH}} \), within the range \(-1 \leq a \leq 1\), where \( J_{\text{BH}} \) is the angular momentum of the SMBH, and \( M_{\text{BH}} \) is the mass of the SMBH. \( a = 0 \) represents a black hole that is not spinning and \( a = 1 \) or \( a = -1 \) represents a maximally spinning black hole.

The SMBH mass grows by gas accretion. To calculate the SMBH spin, \( a' \) after an accretion episode, we use the expression in Bardeen (1970):

\[
a' = \frac{1}{3} \sqrt{\hat{r}_{\text{los},i}} \frac{M_{\text{BH},i}}{M_{\text{BH},f}} \left( 4 - 2 \hat{r}_{\text{los},i} \left( \frac{M_{\text{BH},i}}{M_{\text{BH},f}} \right)^2 - 2 \right)^{1/2},
\]

where \( \hat{r}_{\text{los},i} \) is the radius of the last stable circular orbit in units of the gravitational radius, \( R_G = GM_{\text{BH}}/c^2 \), before the accretion event, \( M_{\text{BH},i} \) is the SMBH mass before the accretion event and \( M_{\text{BH},f} \) is the SMBH mass after the accretion event.

The latter quantities are related to each other by:

\[
M_{\text{BH},f} = M_{\text{BH},i} + (1 - \epsilon_{\text{TD}}) \Delta M,
\]

where \( \Delta M \) is the mass accreted in this accretion event and \( \epsilon_{\text{TD}} \), the radiative accretion efficiency, is given by:

\[
\epsilon_{\text{TD}} = 1 - \left( \frac{2}{3\hat{r}_{\text{los}}} \right)^{1/2}.
\]

\( \hat{r}_{\text{los}} \) is calculated from the spin \( a \), as in Bardeen et al. (1972):

\[
\hat{r}_{\text{los}} = 3 + Z_2 \mp \sqrt{(3 - Z_1)(3 + Z_1 + 2Z_2)},
\]

with the minus sign for \( a > 0 \) and the positive sign for \( a < 0 \). The functions \( Z_1 \) and \( Z_2 \) are given by:

\[
Z_1 = 1 + (1 - |a|^2)^{1/3}(1 + |a|)^{1/3} + (1 - |a|)^{1/3},
\]

\[
Z_2 = \sqrt{3|a|^2 + Z_1^2}.
\]

We consider the accretion disc in three separate parts - an outer disc for the accretion disc at radii greater than the inner radius, \( R_{\text{in}} \), an inner disc for radii less than the inner radius and a warped disc for radii less than the inner radius, \( R_{\text{warp}} \), as shown in Figure 1. The SMBH has an angular momentum \( J_{\text{BH}} \), and the angular momentum of the disc within the inner radius has an angular momentum, \( J_{\text{in}} \), which may not be in the same direction. If that is the case, the spinning black hole induces a Lense-Thirring precession in the misaligned disc elements. Because the precession rate falls off as \( R^{-3} \), at smaller radii the black hole angular momentum and the accretion disc angular momentum vectors will become aligned or anti-aligned, whereas at sufficiently large radii there will still be a misalignment (Bardeen & Petterson 1975). The transition between these two regions occurs at the so-called ‘warp radius’. The angular momentum of the disc within the warp radius is \( J_{\text{warp}} \), which points in the same direction as \( J_{\text{in}} \) initially but with a different magnitude to \( J_{\text{in}} \). When the warp mass is consumed by the SMBH, \( J_{\text{BH}} \) aligns with \( J_{\text{tot}} = J_{\text{BH}} + J_{\text{warp}} \) and \( J_{\text{warp}} \) either anti-aligns or aligns with \( J_{\text{BH}} \) (King et al. 2005). As more gas is accreted, \( J_{\text{BH}} \) eventually aligns with the rest of the inner disc, as the gas in the inner disc is consumed.

There are two proposed modes of how the accretion proceeds once the inner disc has been consumed. In the ‘prolonged mode’ accretion scenario, the angular momentum of the new inner disc is in the same direction as the angular momentum of the outer disc, \( J_{\text{out}} \), but in the ‘chaotic
mode’ accretion scenario introduced in King et al. (2008), the orientation of the angular momentum of the new inner disc is randomised with respect to the angular momentum of the outer disc. King et al. (2008) propose that the inner radius is the self-gravity radius of the disc, which we use in this model. The motivation for chaotic mode accretion is twofold. Firstly, the Soltan (1982) argument, a comparison of the quasar luminosity function at high redshift to the black hole mass function in the local Universe, implies a radiative efficiency of SMBH growth of $\epsilon \approx 0.1$, suggesting that SMBHs in the Universe are typically not maximally spinning, as we would expect from an SMBH that has been spun up by the accretion of gas. Secondly, AGN jets seem to be misaligned with their host galaxies (e.g. Kinney et al. 2000; Sajina et al. 2007), suggesting a misaligned accretion of material onto the SMBH.

Accretion continues in this manner until the gas in the outer disc has been consumed. For this analysis, we adopt chaotic mode accretion as our standard choice.

### 2.3 Warped accretion discs

To obtain the warp radius of an accretion disc, we need expressions for the structure of an accretion disc. Shakura & Sunyaev (1973) introduced the ‘$\alpha$-prescription’ to solve the accretion disc equations, where $\alpha$ is given by $\nu = \alpha c_s H$, with $\nu$ being the viscosity, $c_s$ the sound speed and $H$ the disc semi-thickness. In this analysis, we use the solution of Collin-Souffrin & Dumont (1990), in which the accretion disc equations are solved for AGN discs, using this $\alpha$-prescription. We use their solutions for the regime where opacity is dominated by electron scattering and where gas pressure dominates over radiation pressure.

The disc surface density, $\Sigma$, is given by:

$$\Sigma = 6.84 \times 10^5 \text{g cm}^{-2} \alpha_{TD}^{-4/5} m^{3/5} \left( \frac{M_{BH}}{10^6 M_\odot} \right)^{1/8} \left( \frac{R}{R_s} \right)^{-3/5},$$  

(7)

where $\dot{m} = \dot{M}/M_{Edd}$ is the dimensionless mass accretion rate, $\alpha_{TD}$ is the Shakura-Sunyaev viscosity parameter for thin discs (the value we use is given in Table 3), $R$ is the radius from the centre of the disc and $R_s = 2GM_{BH}/c^2$ is the Schwarzschild radius. The disc semi-thickness $H$ is given by:

$$H = \frac{R}{\dot{\nu}} = 1.25 \times 10^{-3} \alpha_{TD}^{-1/10} \dot{m}^{1/5} \left( \frac{M_{BH}}{10^6 M_\odot} \right)^{-1/10} \left( \frac{R}{R_s} \right)^{1/20},$$

(8)

To calculate the Eddington luminosity we use:

$$L_{Edd} = 1.26 \times 10^{38} \left( \frac{M_{BH}}{10^6 M_\odot} \right) \text{ ergs}^{-1},$$

(9)

and calculate the Eddington mass accretion rate from this using a nominal accretion efficiency $\epsilon = 0.1$:

$$\dot{M}_{Edd} = \frac{L_{Edd}}{0.1 c_s^2};$$

(10)

Note equation (8) is different to Fanidakis et al. (2011) equation (25).

We then follow the method of Natarajan & Pringle (1998) and Volonteri et al. (2007) and take the warp radius as the radius at which the timescale for radial diffusion of the warp is on the order of the local Lense-Thirring precession timescale. This then gives an expression for the warp radius$^3$:

$$R_{warp} = \frac{3410 \alpha_{TD}^{-1/2} \dot{m}^{-1/4} \left( \frac{M_{BH}}{10^6 M_\odot} \right)^{1/8} \left( \frac{\nu_{1,2}}{\nu_c} \right)^{-5/8}}{\Sigma(R) R^2 dR},$$

(11)

where $\nu_{1,2}$ are the horizontal and vertical viscosities respectively. For this analysis, $\nu_c = \nu_2$ (e.g. King et al. 2008). The warp mass can then be calculated using:

$$M_{warp} = \int_0^{R_{warp}} 2\pi \Sigma(R) R^2 dR,$$

(12)

to give an expression$^4$:

$$M_{warp} = 1.35 M_\odot \alpha_{TD}^{-4/5} \dot{m}^{3/5} \left( \frac{M_{BH}}{10^6 M_\odot} \right)^{11/5} \left( \frac{R_{warp}}{R_s} \right)^{7/5}.$$

(13)

### 2.4 Self-gravitating discs

In the chaotic mode accretion scenario of King et al. (2008), the inner radius is the self-gravity radius. The self-gravity radius of accretion disc is the radius at which the vertical gravity due to the disc equals that of the central SMBH. The self-gravity condition is different for the two different types of accretion discs: i) physically thin, optically thick, radiatively efficient ‘thin discs’ (Shakura & Sunyaev 1973) and ii) physically thick, optically thin, radiatively inefficient Advection Dominated Accretion Flows (ADAF - see Yuan & Narayan 2014, for a review). For thin discs (where $\dot{m} > \dot{m}_{crit, ADAF}$), the self-gravity condition is (Pringle 1981):

$$M_{sg} = M_{BH} \frac{H}{R};$$

(14)

where $M_{sg}$ is the self-gravity disc mass, $H$ is the disc semi-thickness (equation (8)) and $R$ is the distance from the black hole (same $R$ as in equation (8)). For ADAFs (where $\dot{m} < \dot{m}_{crit, ADAF}$), $H \sim R$, so this self-gravity condition is:

$$M_{sg} = M_{BH}.$$

(15)

Again using the accretion disc solutions of Collin-Souffrin & Dumont (1990), we derive an expression for the self-gravity radius for thin discs$^3$:

$$R_{sg} = 4790 \alpha_{TD}^{14/27} \dot{m}^{-8/27} \left( \frac{M_{BH}}{10^6 M_\odot} \right)^{-26/27},$$

(16)

Note equation (11) is different to Fanidakis et al. (2011) equation (15).

Note equation (13) is different to Fanidakis et al. (2011) equation (18).

Note equation (16) is different to Fanidakis et al. (2011) equation (24).

MNRAS 000, 1–31 (2018)
and using an integral similar to equation (12), the self-gravity mass for the thin disc is given by:

\[ M_{\text{sg}} = 1.35 M_\odot \alpha_{TD}^{-4/5} \frac{m_3}{m_5} \left( \frac{M_{\text{BH}}}{10^8 M_\odot} \right)^{11/5} \left( \frac{R_{\text{sg}}}{R_8} \right)^{7/5}. \] (17)

### 2.5 Numerical procedure for SMBH accretion

The numerical procedure for the accretion of gas onto the SMBH is as follows. The angular momentum of the SMBH, \( J_{\text{BH}} \), and the angular momentum of the inner disc, \( J_{in} \), are given a random angle in the range \([0, \pi]\) radians between them. The angular momentum vectors change, with \( J_{BH} \) and \( J_{warp} \) aligning and aligning/anti-aligning with \( J_{out} \) respectively. The magnitude of \( J_{BH} \) remains constant, but the magnitude of \( J_{warp} \) changes. This is treated as happening before the mass consumption starts.

We calculate the angular momentum of the material within the warped disc as \( J_{warp} = M_{warp} \sqrt{GM_{BH}R_{warp}} \) and the angular momentum of the black hole, \( J_{BH} = 2^{-1/2} M_{BH} a \sqrt{GM_{BH}R_8} \). Then the ratio of these two quantities is:

\[ \frac{J_{warp}}{2J_{BH}} = \frac{M_{warp}}{2\sqrt{M_{BH}}} \left( \frac{R_{warp}}{R_8} \right)^{1/2}. \] (18)

Whether or not \( J_{warp} \) and \( J_{BH} \) align or anti-align with each other depends on this ratio and on the angle between \( J_{BH} \) and \( J_{in} \) before the accretion event, \( \theta_i \). If \( \cos \theta_i > -J_{warp}/2J_{BH}, J_{warp} \) and \( J_{BH} \) become aligned (prograde accretion), whereas if \( \cos \theta_i < -J_{warp}/2J_{BH}, J_{warp} \) and \( J_{BH} \) become anti-aligned (retrograde accretion). The angle between \( J_{in} \) and \( J_{BH} \) after each accretion event, \( \theta_f \), is:

\[ \cos \theta_f = \frac{J_{warp} + J_{BH}\cos \theta_i}{\sqrt{J_{BH}^2 + J_{warp}^2 + 2J_{warp}J_{BH}\cos \theta_i}}. \] (19)

The model then considers another warp mass of gas with a new \( J_{warp} \) pointing in the same direction as the inner disc and the same process happens again. This repeated process has the effect that \( J_{BH} \) gradually aligns with the angular momentum of the inner accretion disc.

In the prolonged mode, this process keeps happening until all the gas in the outer disc has been consumed whereas in the chaotic mode, once a self-gravity mass of gas has been consumed, the angle between \( J_{in} \) and \( J_{out} \) is randomised. If the mass is being accreted in increments of the self-gravity mass then the ratio of angular momenta is given by:

\[ \frac{J_{in}}{2J_{BH}} = \frac{M_{\text{sg}}}{\sqrt{2m_3 M_{\text{BH}}}} \left( \min\left(\frac{R_{warp}}{R_8}, \frac{R_{\text{sg}}}{R_8}\right) \right)^{1/2}. \] (20)

Here we present predictions for the case in which mass is accreted in increments of the self-gravity mass and assume the chaotic mode of accretion. The output spin distribution of the SMBHs is the same if we use increments of the self-gravity mass or the warp mass and so we use increments of the self-gravity mass as it is less computationally expensive.

In the future we plan a more thorough analysis of the difference between accreting in increments of self-gravity mass versus increments of warp mass. The AGN luminosities are not affected by our choice in this area as they depend on the accreted mass and the SMBH spin as we see in Section 2.7.

### 2.6 Spinup by SMBH mergers

The other way by which SMBHs can change their spin is by merging with another SMBH. The spin value of the subsequent SMBH will depend on the spin values of the two SMBHs, but also on the angular momentum of their binary orbit. To determine the final spin value, \( a_f \), we use the expressions obtained from numerical simulations of BH-BH mergers in Rezzolla et al. (2008):

\[ a_f = \frac{1}{(1+q)^2} \left( |a_1|^2 + |a_2|^2 q^4 + 2|a_1||a_2|q^2 \cos \phi + 2(|a_1| \cos \theta + |a_2|q^2 \cos \xi) |l|q + |l|^2 q^2 \right)^{1/2}. \] (21)

where the magnitude of \( l \), the orbital angular momentum, is given by:

\[ |l| = \frac{s_4}{(1+q^2)^2} \left( |a_1|^2 + |a_2|^2 q^4 + 2|a_1||a_2|q^2 \cos \phi + \frac{s_5 \mu + s_0 + 2}{1+q^2} \right) |(a_1| \cos \theta + |a_2|q^2 \cos \xi) + 2\sqrt{3} t_2 \mu + t_3 \mu^2. \] (22)

Here, \( s_4 = -0.129 \), \( s_5 = -0.384 \), \( t_0 = -2.686 \), \( t_2 = -3.454 \), \( t_3 = 2.353 \) are values obtained in Rezzolla et al. (2008). \( a_{1,2} \) are the spins of the SMBHs, \( q \) is the mass ratio \( M_1/M_2 \), \( \mu \) is the symmetric mass ratio \( q/(q+1)^2 \). \( M_1 \) and \( M_2 \) are chosen such that \( q/\text{eq}1 \). The angles \( \phi, \theta \) and \( \xi \) are given by:

\[ \cos \phi = \tilde{a}_1 \cdot \tilde{a}_2, \] (23)

\[ \cos \theta = \tilde{a}_1 \cdot \tilde{l}, \] (24)

\[ \cos \xi = \tilde{a}_2 \cdot \tilde{l}. \] (25)

When we consider two SMBHs merging, we calculate the angles between the three different vectors by randomly selecting directions uniformly over the surface of a sphere. This prescription makes the assumption that the radiation of gravitational waves does not affect the direction of the orbital angular momentum, and we also assume that the mass lost to gravitational radiation is negligible.

### 2.7 AGN bolometric luminosities

These accreted gas masses then allow us to calculate a radiative bolometric luminosity for an accreting SMBH. In the starburst mode, we assume that during an accretion episode the accretion rate is constant over a time \( t_{\text{bulge}} \), where \( t_{\text{bulge}} \) is the dynamical timescale of the bulge and \( f_q \) is a

---

Note equation (17) is different to Fanidakis et al. (2011) equation (26).
free parameter, given in Table 3. Therefore, the mass accretion rate is given by:

\[ \dot{M} = \frac{M_{\text{acc}}}{f_b \text{bolage}} \]  

(26)

where \( M_{\text{acc}} \) is the total accreted mass, which is equal to \( f_{\text{BH}} M_\odot \) bolage, where \( f_{\text{BH}} \) is a free parameter (c.f. Lacey et al. 2016) and \( M_\odot \text{bolage} \) is the mass of stars formed in the starburst. Results from Malbon et al. (2007) suggest that the results are similar is the accretion timescale is a top hat function or an exponential. In the hot halo mode, which is only active when AGN feedback is active, the mass accretion rate is the mass accretion rate corresponding to that which suppresses cooling:

\[ \dot{M} = \frac{L_{\text{cool}}}{\epsilon_{\text{heat}} c^2} \]  

(27)

where \( L_{\text{cool}} \) is the radiative cooling luminosity, and \( \epsilon_{\text{heat}} \) is a free parameter (c.f. Lacey et al. 2016). We then calculate a thin disc bolometric luminosity using:

\[ L_{\text{bol,TD}} = \epsilon_{\text{TD}} \dot{M} c^2, \]  

(28)

where we calculate the accretion efficiency \( \epsilon_{\text{TD}} \) for the thin disc case using equation (3). However, the accretion efficiency is not the same for all regimes of accretion flow. As well as the thin disc and the ADAF, which will have different accretion efficiencies, there are also AGNs accreting above the Eddington accretion rate. Such objects are understood to be advection dominated and have optically thin flows (Abramowicz et al. 1988).

For the ADAF regime we use the expressions for bolometric luminosity from Mahadevan (1997). There are two cases within this regime. For lower accretion rate ADAFs (\( \dot{m} < m_{\text{crit,visc}} \)), electron heating is dominated by viscous heating, whereas for higher accretion rate ADAFs (\( m_{\text{crit,visc}} < \dot{m} < m_{\text{crit,ADAF}} \)), the ion-electron heating dominates the electron heating. In the super-Eddington regime, the radiative efficiency is lower than the corresponding thin disc radiative efficiency, and so a super-Eddington luminosity supression is introduced (Shakura & Sunyaev 1973). This expression includes a free parameter, \( \eta_{\text{Edd}} \), the value for which is given in Table 3.

Hence, the bolometric luminosities in the model are given by the following. For \( \dot{m} < m_{\text{crit,visc}} \):

\[ L_{\text{bol}} = 0.0002 \epsilon_{\text{TD}} \dot{M} c^2 \left( \frac{\delta_{\text{ADAF}}}{0.0005} \right) \left( \frac{1 - \beta}{0.5} \right) \left( \frac{6}{\tau_{\text{bol}}} \right), \]  

(29)

for \( m_{\text{crit,visc}} < \dot{m} < m_{\text{crit,ADAF}} \):

\[ L_{\text{bol}} = 0.2 \epsilon_{\text{TD}} \dot{M} c^2 \left( \frac{\dot{m}}{m_{\text{ADAF}}^2} \right) \left( \frac{\beta}{0.5} \right) \left( \frac{6}{\tau_{\text{bol}}} \right), \]  

(30)

for \( m_{\text{crit,ADAF}} < \dot{m} < \eta_{\text{Edd}} \):

\[ L_{\text{bol}} = \epsilon_{\text{TD}} \dot{M} c^2, \]  

(31)

Note that the coefficients of the ADAF luminosities are derived in Mahadevan (1997) and not free parameters.

and for \( \dot{m} > \eta_{\text{Edd}} \):

\[ L_{\text{bol}} = \eta_{\text{Edd}} (1 + \ln(\dot{m}/\eta_{\text{Edd}})) L_{\text{Edd}}. \]  

(32)

where \( \delta_{\text{ADAF}} \) is the viscosity parameter in the ADAF regime (the value is given in Table 3). \( f_b \) and \( \eta_{\text{Edd}} \) are free parameters that we calibrate on observed AGN luminosity functions, as described in Section 4.1. \( \delta_{\text{ADAF}} \) is the fraction of viscous energy transferred to the electrons (the value is given in Table 3), for this study we adopt a value \( \delta_{\text{ADAF}} = 0.2 \), more in line with observational (Yuan et al. 2003; Liu & Wu 2013) and theoretical (Sharma et al. 2007) constraints, as opposed to the value of \( \delta_{\text{ADAF}} = 2000^{-1} \) adopted in Fanidakis et al. (2012). The current consensus for the value of \( \delta_{\text{ADAF}} \) is a value between 0.1 and 0.5, (c.f. Yuan & Narayan 2014). Changing the value of \( \delta_{\text{ADAF}} \) makes no discernable difference to the luminosity functions shown in this paper. \( \beta \) is the ratio of gas pressure to total pressure (total pressure being the sum of gas pressure and magnetic pressure). Following Fanidakis et al. (2012), we use the relation \( \beta = 1 - 0.1 \delta_{\text{ADAF}} / 0.55 \), which is based on MHD simulations in Hawley et al. (1995).

The boundary between the two ADAF regimes is:

\[ m_{\text{crit,visc}} = 0.001 \left( \frac{\delta_{\text{ADAF}}}{0.0005} \right) \left( \frac{1 - \beta}{\beta} \right) \left( \frac{\alpha_{\text{ADAF}}^2}{\alpha_{\text{Edd}}^2} \right), \]  

(33)

which is a value chosen so that \( L_{\text{bol}} \) is continuous for ADAF luminosities. The boundary between the ADAF and thin disc regimes is assumed to be \( m_{\text{crit,ADAF}} = 0.01 \) (Yuan & Narayan 2014).

### 2.8 Converting from bolometric to optical and X-ray AGN luminosities

To convert from bolometric luminosity to luminosities in other wavebands we use the bolometric corrections derived from the empirical AGN SED in Marconi et al. (2004). The rest-frame bolometric corrections calculated from this SED are:

\[ \log_{10}(L_{\text{HX}}/L_{\text{bol}}) = -1.54 - 0.24 \mathcal{L} - 0.012 \mathcal{L}^2 + 0.0015 \mathcal{L}^3, \]  

(34)

\[ \log_{10}(L_{\text{SX}}/L_{\text{bol}}) = -1.65 - 0.22 \mathcal{L} - 0.012 \mathcal{L}^2 + 0.0015 \mathcal{L}^3, \]  

(35)

\[ \log_{10}(\nu_B L_{\nu B}/L_{\text{bol}}) = -0.80 + 0.067 \mathcal{L} - 0.017 \mathcal{L}^2 + 0.0023 \mathcal{L}^3, \]  

(36)

where \( \mathcal{L} = \log_{10}(L_{\text{bol}}/10^{42} L_\odot) \), \( L_{\text{HX}} \) is the hard X-ray (2-10 keV) luminosity, \( L_{\text{SX}} \) is the soft X-ray (0.5-2 keV) luminosity, \( \nu_B \) is the frequency of the centre of the B-band, \( c/4400 \AA \), and \( L_{\nu B} \) is the luminosity in the B-band.

To calculate B-band magnitudes we use the expression:

\[ M_{B,AB} = -11.33 - 2.5 \log_{10} \left( \frac{\nu_B L_{\nu B}}{10^{40} \text{ergs}^{-1}} \right), \]  

(37)

\[ \log_{10}(L_{\text{HX}}/L_{\text{bol}}) = -1.54 - 0.24 \mathcal{L} - 0.012 \mathcal{L}^2 + 0.0015 \mathcal{L}^3, \]  

Note equations (34) and (35) are corrected from Fanidakis et al. (2012) equation (10).

Note equation (37) is different to Fanidakis et al. (2012) equation (13).
for magnitudes in the AB system, from the definition of AB magnitudes (Oke & Gunn 1983). Using the Marconi et al. (2004) SED template, we convert from rest-frame B-band magnitudes to rest-frame 1500 Å band magnitudes using a relation similar to those in Appendix B to give:

$$M_{1500,AB} = M_{B,AB} + 0.514.$$  

(38)

It is worth noting that the AGN SED we use is based on observations of quasars and therefore probably most appropriate for AGN in the thin disc regime. For $z > 6$ and for the luminosities that we are considering, the AGN are in the thin disc or super-Eddington regime, so this SED is appropriate, although in future work we plan to include a wider variety of SED templates for AGNs in different accretion regimes.

2.9 AGN obscuration and visible fractions

AGN are understood to be surrounded by a dusty torus, which causes some of the radiation to be absorbed along some sightlines, and re-emitted at longer wavelengths. Therefore, the population of AGN generated by the model will contain some obscured objects and some unobscured objects. For simplicity, we assume that at a given wavelength, AGN are either completely obscured or completely unobscured. The effect of obscuration can therefore be expressed as a visible fraction, which is the proportion of objects that are visible at a certain luminosity and redshift, and in a certain waveband.

The fraction of obscured objects in the hard X-ray band is also thought to be small, so for this work we assume that there is no obscuration at hard X-ray wavelengths. There is a population of so-called ‘Compton-thick’ AGNs for which the column density of neutral hydrogen exceeds the unit optical depth corresponding to the Thomson cross section ($N_H \approx 1.5 \times 10^{24} \text{cm}^{-2}$). Such objects are difficult to detect, even at hard X-ray wavelengths. The number of such objects is thought to be small, so we ignore their contribution for this work.

We calculate the visible fractions in the soft X-ray and optical using three observationally determined empirical relations from the literature, and also two more introduced in this work.

(i) The visible fraction of Hasinger (2008) is:

$$f_{\text{vis}} = 1 + 0.281 \left[ \log_{10} \left( \frac{L_{\text{HX}}}{10^{48} \text{ergs}^{-1}} \right) \right] - A(z),$$  

(39)

where

$$A(z) = 0.279 (1 + z)^{0.62}.$$  

(40)

$L_{\text{HX}}$ is the hard X-ray luminosity in the observer frame and $z$ is the redshift. The redshift dependence of the visible fraction in this model saturates at $z \geq 2.06$ and the visible fraction is not allowed to have values below 0 or above 1.

This model and others we use from observational studies assume a slightly different cosmology from the one used in the P-Millennium, for simplicity we ignore the effect of this here.

Because the observational data on which the model is based only extend to $z = 2$, for $z > 2$, we calculate it using $L_{\text{HX}}$ as the rest-frame hard X-ray band at $z = 2$, i.e. 6-30 keV. For this obscuration model, if an object is obscured at soft X-ray wavelengths, then it is also assumed to be obscured at optical wavelengths.

(ii) Hopkins et al. (2007) derive a visible fraction of the form:

$$f_{\text{vis}} = f_{\text{abs}} \left( \frac{L_{\text{bol}}}{10^{46} \text{ergs}^{-1}} \right)^{\beta},$$  

(41)

where $f_{\text{abs}}$ and $\beta$ are constants for each band. For the B-band, $[f_{\text{abs}}, \beta]$ is $[0.260, 0.082]$ and for the soft X-ray band, $[f_{\text{abs}}, \beta]$ is $[0.609, 0.063]$. This model does not require a high redshift correction, as it depends on bolometric luminosity.

(iii) Aird et al. (2015) observationally determine a visible fraction for soft X-rays of the form:

$$f_{\text{vis}} = \frac{\phi_{\text{unabs}}}{\phi_{\text{unabs}} + \phi_{\text{abs}}},$$  

(42)

where $\phi$, the number density of absorbed/unabsorbed sources, is given by:

$$\phi = \frac{K}{(L_{\text{HX}}/L_{\text{bol}})^{\gamma_1} + (L_{\text{HX}}/L_{\text{bol}})^{\gamma_2}},$$  

(43)

where the constants for both cases are given in Table 2. As for the Hasinger (2008) obscuration model, if the object is obscured at soft X-ray wavelengths, then we assume that it is also obscured at optical wavelengths. For this obscuration model, we correct for high redshift such that for $z > 3$, the $L_{\text{HX}}$ band is the rest-frame band for $z = 3$.

(iv) We also use visible fractions that are modified versions of Hopkins et al. (2007). These visible fractions also depend solely on $L_{\text{bol}}$, but with different coefficients. These coefficients were derived from constructing a bolometric luminosity function from the observational datasets. We used the Marconi et al. (2004) bolometric corrections and selected coefficients for the visible fraction to reduce the scatter in the resultant bolometric luminosity function. This is discussed in Appendix C. The first of these, the ‘low-z modified Hopkins’, (LZMH) visible fraction for rest-frame 1500 Å has the form:

$$f_{\text{vis},\text{LZMH}} = 0.15 \left( \frac{L_{\text{bol}}}{10^{46} \text{ergs}^{-1}} \right)^{-0.1},$$  

(44)

and for the soft X-ray band it has the form:

$$f_{\text{vis},\text{LZMH}} = 0.4 \left( \frac{L_{\text{bol}}}{10^{46} \text{ergs}^{-1}} \right)^{0.1}.$$  

(45)

(v) The second of these modified Hopkins visible fractions, the ‘$z = 6$ modified Hopkins’ (Z6MH) visible fraction for rest-frame 1500 Å was derived by fitting the model to the $z = 6$ optical luminosity function. This visible fraction is:

$$f_{\text{vis},\text{Z6MH}} = 0.04 \left( \frac{L_{\text{bol}}}{10^{46} \text{ergs}^{-1}} \right)^{0.0} = 0.04.$$  

(46)

There is not an equivalent Z6MH visible fraction for soft X-rays, as by $z = 6$ the rest-frame is at hard X-ray wavelengths where we assume no obscuration effects.
2.10 Calculating model AGN luminosity functions

Typically when one constructs a luminosity function from a simulation, only the AGN that are switched on at each snapshot are included. However, if one does this, rarer objects with higher luminosities but which are only active for a short time are not sampled well. To probe the luminosity function for such objects, we must average over a time window, \( t_{\text{window}} \). The time window cannot be too large, as we may miss the effect of multiple starbursts within the time window, because the simulation only outputs information on the most recent starburst. We select a time window for which the luminosity function appears converged. For the predictions here we set \( t_{\text{window}} = t_{\text{snapshot}}/10 \), where \( t_{\text{snapshot}} \) is the age of the Universe at that redshift.

Each object is assigned a weight, \( w \), given by:

\[
w = \frac{t_Q}{t_{\text{window}}},
\]

where \( t_Q = q_{\text{bulge}} \) is the lifetime of the quasar event as in Section 2.7. This weight is then assigned to the number densities which then allows us to include higher luminosity events at lower number densities in the luminosity function.

3 SMBH MASSES, ACCRETION RATES AND SPINS

3.1 Black hole masses

We start by showing some basic predictions from the new model for SMBH masses, accretion rates and spins. In the left panel of Figure 2 we show the black hole mass function at \( z = 0 \) predicted by our model compared to observational estimates. The observations use indirect methods to estimate the black hole mass function, because of the lack of a large sample of galaxies with dynamically measured black hole masses. In Marconi et al. (2004) and Shankar et al. (2004, 2009) galaxy luminosity/velocity dispersion functions are combined with relations between black hole mass and host galaxy properties to estimate black hole mass functions. The predictions of the model fit well to the observational estimates within the observational errors, especially given that there will also be uncertainties on the black hole mass measurements and given the discrepancies between the observational estimates. The former means the predictions could still be consistent at the high mass end.

The evolution of the black hole mass function for \( 0 < z < 12 \) is shown in the right panel of Figure 2. Most of the SMBH mass is formed by \( z \approx 2 \), as the mass density of black holes is dominated by objects around the knee of the black hole mass function, and this knee is place by \( z \approx 2 \). The dominant fuelling mechanism for growing the black hole mass density across all redshifts is gas accretion in starbursts triggered by disc instabilities, and disc instabilities play an important role in shaping the black hole mass function for \( M_{\text{BH}} < 10^7 M_\odot \). However, SMBH mergers are more important for determining the shape of the black hole mass function for \( M_{\text{BH}} > 10^7 M_\odot \), as they are the mechanism by which the largest SMBHs are formed. AGN feedback also plays an important role in shaping the black hole mass function at this high mass end, by suppressing gas cooling and so slowing down the rate at which the SMBHs grow by cold gas accretion.

In the left panel of Figure 3 we show the predicted SMBH mass versus bulge mass relation compared to observational data from McConnell & Ma (2013). The predictions follow the observations well, with the scatter decreasing towards higher masses. BH-BH mergers contribute towards this decrease in scatter, as seen in Jalkanen & Macciò (2011), although they are not the only contributing mechanism, with AGN feedback also affecting the scatter at the high mass end.

In the middle panel of Figure 3, we show the evolution of the SMBH mass versus bulge stellar mass relation for \( 0 < z < 6 \), showing the scatter of the distribution for \( z = 0 \) and \( z = 6 \). As we go to higher redshift, the ratio \( M_{\text{BH}} / M_{\text{bulge}} \) increases, as also seen in the observations (e.g. Peng et al. 2006). The ratio \( M_{\text{BH}} / M_{\text{bulge}} \) reflects the mechanism by which these two galaxy components form. At higher redshift, bulges grow by starbursts, which also feeds the growth of SMBHs and so the distribution of the ratio \( M_{\text{BH}} / M_{\text{bulge}} \) peaks at \( f_{\text{BH}} \) (the fraction of mass accreted onto a black hole in a starburst event), with some scatter caused by mergers. At lower redshift the ratio \( M_{\text{BH}} / M_{\text{bulge}} \) decreases, as mergers cause bulges to form from discs, but without growing the SMBHs. We also note how the scatter of the SMBH mass vs. bulge mass relation is lower at \( z = 6 \) than at \( z = 0 \) for all masses - by \( z = 0 \) galaxies have had a more varied formation history compared to the \( z = 6 \) population.

In the right panel of Figure 3 we show the evolution of the SMBH mass versus galaxy stellar mass relation with redshift for the redshift range \( 0 < z < 6 \). Galaxies of larger stellar mass and the largest SMBHs form at late times, and the slope of the relation at the low mass end is steeper at later times. At the low mass end, the ratio \( M_{\text{BH}} / M_{\star} \) decreases with time as stellar mass builds up around the peak of star formation at \( z \approx 2 \). This evolution slows down at \( z < 1 \). At the high mass end, the stellar mass and SMBH mass stay on the same relation independent of redshift. It is in this regime that the AGN feedback is operational: in
Table 3. The values for the SMBH/AGN free parameters in the model. The upper part of the table shows parameters where the values adopted are from other studies, whereas the lower part of the table gives parameters which have been calibrated on the luminosity functions shown here.

| Parameter | Fanidakis et al. (2012) | Adopted here | Significance |
|-----------|------------------------|--------------|--------------|
| $\alpha_{\text{ADAF}}$ | 0.087 | 0.1 | Shakura & Sunyaev (1973) viscosity parameter for ADAFs |
| $\alpha_{\text{TD}}$ | 0.087 | 0.1 | Shakura & Sunyaev (1973) viscosity parameter for TDs |
| $\delta_{\text{ADAF}}$ | $2000^{-1}$ | 0.2 | Fraction of viscous energy transferred to electrons in ADAF |
| $\dot{m}_{\text{crit},\text{ADAF}}$ | 0.01 | 0.01 | Boundary between thin disc and ADAF accretion |
| $\eta_{\text{Edd}}$ | 4 | 4 | Super-Eddington suppression factor |
| $f_\eta$ | 10 | 10 | Ratio of lifetime of quasar event and bulge dynamical timescale |

Figure 2. The black hole mass function. *Left panel:* the predicted black hole mass function at $z = 0$ compared to observational estimates by Marconi et al. (2004); Shankar et al. (2004, 2009). *Right panel:* the evolution of the black hole mass function over the range $0 < z < 12$.

In our model we use the AGN feedback prescription of Bower et al. (2006) in which AGN feedback is only active where the hot gas halo is undergoing ‘quasistatic’ (slow) cooling. This has the effect that AGN feedback is only active for haloes of mass above $\sim 10^{12} M_\odot$. The relation between SMBH mass and stellar mass at this high mass end is caused by both AGN feedback and mergers, with neither mechanism dominant in establishing this relation.

3.2 Black hole accretion rates

In Figure 4 we show the black hole mass accretion rate distribution showing its evolution with redshift and split by fuelling modes: the hot halo mode, starbursts triggered by mergers and starbursts triggered by disc instabilities (see Section 2). The hot halo mode becomes more dominant at later times, because the hot halo mode requires long cooling times, and hence it occurs for massive haloes, and because dark matter haloes grow hierarchically, these large haloes only form at later times. The contribution from starbursts triggered by galaxy mergers peaks at $z \approx 2$. Starbursts triggered by mergers peak at a low mass accretion rate, as seen in Figure 4, albeit with a tail that extends to high $\dot{M}$. The peak at $\dot{M} \sim 10^{-7} M_\odot/yr$ is mostly due to minor mergers with mass ratios $0.05 < M_2/M_1 < 0.3$ (mergers with mass ratios in this range cause about three quarters of the starbursts at this mass accretion rate)\footnote{Note that a mass ratio of 0.05 is the lower threshold for starburst triggering in galaxy mergers (Lacey et al. 2016).}. The contribution from starbursts triggered by disc instabilities increases as the redshift increases. Starbursts triggered by mergers typically have lower $\dot{M}$ values than starbursts triggered by disc instabilities. There are two reasons for this. Firstly, the average stellar mass formed by bursts triggered by disc instabilities is higher than for bursts triggered by mergers, and this occurs because the average cold gas mass is higher for galaxies in which bursts triggered by disc instabilities occur. Secondly, the average bulge dynamical timescale for
High redshift AGN predictions

Figure 3. Left panel: the predicted SMBH mass versus bulge stellar mass relation at $z = 0$ compared to observational data from McConnell & Ma (2013). The line represents the median of the predicted SMBH mass in bins of bulge mass and the shading denotes the 10-90 percentiles of the predicted distribution. Middle panel: the evolution of the median of the SMBH mass versus bulge mass relation with redshift for $z = 0, 0.5, 1, 2, 4, 6$. As in the left panel, the grey shaded band is the 10-90 percentiles of the distribution for $z = 0$ and the purple dashed lines are the 10-90 percentiles of the distribution for $z = 6$. Right panel: the evolution of the median of the SMBH mass versus galaxy stellar mass relation, with the lines representing the same redshifts as the middle panel as indicated by the legend.

Figure 4. The distribution of black hole mass accretion rates for different redshifts (black solid line) split by contributions from hot halo mode (red dashed line), starbursts triggered by mergers (light blue solid line) and starbursts triggered by disc instabilities (dark blue dotted line). We have only selected black holes residing in galaxies of stellar mass, $M_\star > 10^6 M_\odot$, which is above the completeness limit of the simulation.
starbursts triggered by disc instabilities is smaller than for
together due to the average bulge size being smaller for starbursts triggered by disc instabilities. The combination of these effects accounts for the lack of star bursts triggered by disc instabilities at the very lowest \( \dot{M} \) values. The galaxies that host such starburst events would be below the mass at which the simulation is complete.

### 3.3 Black hole spins

In Figure 5, we show the SMBH spin distribution output from the model for both the prolonged and chaotic accretion modes. Note that \( \alpha \) here represents the magnitude of the spin. The low mass end of the spin distribution (6 < \( \log_{10}(M_{\text{BH}}/M_\odot) < 8 \)) is dominated by accretion spinup whereas the high mass end (8 < \( \log_{10}(M_{\text{BH}}/M_\odot) < 10 \)) is dominated by merger spinup. For prolonged mode accretion, the coherent accretion spinup means that SMBHs quickly reach their maximum spin value, giving rise to a population of maximally spinning SMBHs at low mass. At high masses, the average spin value is lower because of SMBH mergers. This is because even if two maximally spinning SMBHs merge, the result is typically a SMBH with a lower spin value. For chaotic mode accretion, the accretion direction is constantly changing and so the accretion spinup leads to SMBHs with lower spin values (\( \bar{\alpha} \approx 0.4 \)), compared to prolonged accretion. The spin values are not zero in the chaotic mode, as one may be tempted to expect, because the accretion spinup is more efficient if the accretion disc and SMBH spin are in the same direction compared to the case of anti-alignment (King et al. 2008). The mean value of the SMBH spin decreases with increasing black hole mass at this low mass end, for the chaotic mode accretion spinup case as also reported in King et al. (2008). At the high mass end, the increase in average spin at \( M_{\text{BH}} \sim 10^5 M_\odot \) is due to spinup by BH mergers. Two slowly spinning SMBHs typically form a higher spin SMBH when they merge, due to the angular momentum of the orbit between them providing angular momentum for the spin of the resultant SMBH.

One of the conclusions of Fanidakis et al. (2011) was that for chaotic mode accretion, smaller SMBHs will have lower spin values (\( \bar{\alpha} \approx 0.15 \)) whereas larger SMBHs will have higher spin values (\( \bar{\alpha} \approx 0.7 - 0.8 \)). Our new analysis predicts that for chaotic mode accretion SMBHs will moderately spin values, \( \bar{\alpha} \approx 0.4 \), yielding radiative accretion efficiencies of \( \epsilon \approx 0.075 \), not too dissimilar from the value of \( \epsilon \approx 0.1 \) required by the Soltan (1982) argument. However, the average radiative accretion efficiency implied by prolonged mode accretion is \( \epsilon \approx 0.4 \), in tension with the Soltan (1982) argument.

The chaotic mode spin distribution is different to that in Fanidakis et al. (2011) because the equations for SMBH spinup by gas accretion have changed from that paper (causing higher spin values at the low SMBH mass end) and because the directions for the spinup due to SMBH mergers are sampled from the surface of a sphere as opposed to the circumference of a circle, leading to lower spin values at the high SMBH mass end.

### 4 EVOLUTION OF THE AGN LUMINOSITY FUNCTION AT Z < 6

#### 4.1 Bolometric luminosity function

First, we present the predicted bolometric luminosity function compared to our observationally estimated bolometric luminosity function constructed from the observations. This observationally estimated bolometric luminosity function described in Appendix C is compared to other observational estimates from Hopkins et al. (2007).

The model for SMBH evolution and AGN luminosity also involves some free parameters additional to those in the galaxy formation model, as shown in Table 3. We have calibrated the values of \( f_L \) and \( \eta_{\text{MDR}} \), and found that the best-fitting values are those adopted in Fanidakis et al. (2012). We show the effect of varying these parameters in Figures D1 and D2. We also slightly adjust the values of \( \alpha_{\text{ADAF}} \) and \( \epsilon_{\text{TDE}} \) from 0.087 to 0.1. This is for simplicity and to keep the values in line with MHD simulations (e.g., Penna et al. 2013). The value of \( \delta_{\text{ADAF}} \) has been updated from Fanidakis et al. (2012) (c.f. Section 2.7).

We show the predicted bolometric luminosity function compared to observational estimates first, in Figure 6, as it is the simplest AGN luminosity output from the model, as opposed to the AGN luminosities in other bands that are calculated via a bolometric correction based on a template SED. The predictions (the black line is the sum of the contributions from all accretion modes) compare well to the observational bolometric luminosity function across the range of redshifts and for the luminosities shown. Exceptions include the faint end at high redshift where the model underpredicts the observations by 0.5 dex for \( L_{\text{bol}} < 10^{44}\text{ergs}^{-1} \), for \( z > 4 \), and the faint end at low redshift where the model underpredicts the observations for \( L_{\text{bol}} < 10^{45}\text{ergs}^{-1} \) and \( z < 0.5 \) by 0.5 dex. The underpredictions at the faint end at low redshift may be because the ADAF radiative accretion efficiency is lower than the thin disc accretion efficiency, leading to lower luminosities (see Figure D5 for a prediction with only a thin disc accretion efficiency for all values of \( \dot{m} \)). Alternatively, this discrepancy could be resolved by assuming an accretion timescale with a dependence on accreted gas mass or black hole mass. In a different model, Shirakata et al. (2018) obtain a better fit to the hard X-ray luminosity function at the low luminosity end at low redshift by doing this. In general, our model is a good match to these observations across a broad range.

We also show in Figure 6 the separate contributions to the AGN luminosity function from ADAFs (\( \dot{m} < \dot{m}_{\text{crit,ADAF}} \)), thin discs (\( \dot{m}_{\text{crit,ADAF}} < \dot{m} < \dot{m}_{\text{Real}} \)) and super-Eddington objects (\( \dot{m} > \dot{m}_{\text{Real}} \)). At low redshift, ADAF dominate the faint end (\( L_{\text{bol}} < 10^{44}\text{ergs}^{-1} \)), thin discs dominate at intermediate luminosities (\( 10^{44}\text{ergs}^{-1} < L_{\text{bol}} < 10^{46}\text{ergs}^{-1} \)) and super-Eddington objects dominate the bright end (\( L_{\text{bol}} > 10^{46}\text{ergs}^{-1} \)). As we go to higher redshift, the ADAF’s contribution to the luminosity function decreases: for \( 0 < z < 2 \) the evolution is not that strong, although the contribution from ADAFs at each luminosity decreases slightly as we increase \( z \) in this range, whereas for \( z > 2 \), the evolution in the ADAF population is pronounced, and the number of ADAFs drops off sharply with increasing redshift. In contrast, the contribution from the thin disc population increases until \( z \approx 2 \), after which it remains ap-
The SMBH spin distribution from the model at $z = 0$ for prolonged and chaotic accretion mode. The line represents the median value of the magnitude of the spin for that SMBH mass, and the shading represents the 10-90 percentiles of the distribution.

Figure 5. The SMBH spin distribution from the model at $z = 0$ for prolonged and chaotic accretion mode. The line represents the median value of the magnitude of the spin for that SMBH mass, and the shading represents the 10-90 percentiles of the distribution.

Approximately constant. At $z < 2$, there are not very many super-Eddington objects and so they make a fairly small contribution to the luminosity function but their contribution increases at $z > 2$. The distribution of super-Eddington objects is bimodal, and for $z < 4$, the higher luminosity peak has a higher number density, while for $z > 4$, the lower luminosity peak has a higher number density. The bimodality is not due to the bimodality in the fuelling modes, as all the super-Eddington objects are fuelled by starbursts triggered by disc instabilities, but it seems to be caused by a bimodality in the bulge stellar mass. We plan to explore this issue in more detail in future work.

In Figure 7 we split the AGN luminosity function by contributions from the hot halo mode, starbursts triggered by mergers and starbursts triggered by disc instabilities. At low redshift ($z < 2$), the faint end is dominated by the hot halo mode, whereas the bright end is dominated by starbursts triggered by disc instabilities. Starbursts triggered by mergers make a small contribution to the AGN bolometric luminosity function at low redshift. Starbursts triggered by disc instabilities typically have higher values of $\dot{M}$ and so higher luminosities compared to starbursts triggered by mergers, which is why they dominate the bright end.

The hot halo mode only operates in the most massive haloes, and so the hot halo mode only begins to significantly contribute to the AGN luminosity function for $z < 3$. The hot halo mode contribution to the luminosity function is in place at $z = 2$ and it changes little afterwards. For $z > 2$, starbursts triggered by disc instabilities dominate the AGN luminosity function, with starbursts from mergers not significantly contributing. This implies that the inclusion of black hole growth via disc instabilities is significant for reproducing AGN luminosity functions at high redshift.

A key aspect of the success of the GAlFORM AGN model is the different channels of black hole growth, particularly the inclusion of disc instability triggered starbursts, that allow a good match to the AGN luminosity functions to be obtained. Other semi-analytic models do not necessarily include disc instabilities, which may explain why they do not reproduce AGN properties particularly well at high redshift (e.g. Bonoli et al. 2009; Menci et al. 2013; Neistein & Netzer 2014; Enoki et al. 2014). The effect of disc instabilities on the AGN predictions at $0 < z < 6$ is shown in Figure D3 and the effect on galaxy properties is shown in Lacey et al. (2016).

We show the effect on the AGN bolometric luminosity function of changing between chaotic and prolonged mode in Figure 8. In the prolonged mode, SMBH spins are higher (see Figure 5), and so this results in a higher radiative accretion efficiency leading to higher bolometric luminosities.

4.2 Luminosity functions in observed bands

We then use the conversions described in Section 2.8 and visible fractions described in Section 2.9 to make predictions for the luminosity function in the rest-frame hard X-ray, soft X-ray and 1500Å bands. We compare our hard X-ray predictions to Ueda et al. (2003, 2014); Aird et al. (2015) in Figure 9. The model is generally in good agreement with the data, particularly in the range $1 < z < 3$. For $L_{\text{H}X} < 10^{44}\text{ergs}^{-1}$ at $z < 0.5$, the model underpredicts the observations by about 0.5 dex, and for $L_{\text{H}X} < 10^{44}\text{ergs}^{-1}$ at $z > 3$, the model overpredicts the observations by about 1 dex.

The former discrepancy corresponds to the model bolometric luminosity function underpredicting the observations in the same redshift and luminosity regime, and the latter also corresponds to the bolometric luminosity function slightly overpredicting the observational estimates in that regime, but may also be influenced by our assumption that there is no obscuration for hard X-ray sources. This assumption may be not valid for the high redshift Universe; more observations are needed to constrain the obscuration effect on hard X-rays.

Our soft X-ray predictions are compared to Hasinger

Note that the shape of the luminosity function changes little between the two models.
Figure 6. The AGN bolometric luminosity function predicted by our model (black line, with grey shading showing the Poisson errorbars) compared to our bolometric luminosity function constructed from the observations. We show the observational data indicating the wavelength of the data that was used to construct that particular point (squares - hard X-ray, triangles - soft X-ray, circles - optical). We split the total bolometric luminosity function by accretion mode into ADAFs (green), thin discs (purple) and super-Eddington objects (grey).

Figure 7. The AGN bolometric luminosity function as Figure 6, but split by the fuelling mode: hot halo mode (red), starbursts triggered by mergers (light blue), starbursts triggered by disc instabilities (dark blue).
Figure 8. The effect of changing between chaotic (blue) and prolonged (red) mode on the AGN bolometric luminosity function at $z = 0.2, 2, 6$.

Figure 9. The hard X-ray luminosity function generated by the model (black line) compared to hard X-ray luminosity function observations from Ueda et al. (2003) (circles), Ueda et al. (2014) (squares) and Aird et al. (2015) (triangles).

et al. (2005), and Aird et al. (2015) in Figure 10. The luminosity function without taking into account obscuration is shown alongside the model with the visible fractions of Hopkins et al. (2007), Hasinger (2008), Aird et al. (2015) and our observationally determined LZMH model. The luminosity functions with different visible fractions are very similar except for $L_{\text{SX}} < 10^{44}\text{ergs}^{-1}$. The LZMH model fits best in the range $1 < z < 2$. At higher redshifts and lower luminosities the visible fraction in the Hasinger (2008) model drops to zero, which causes the corresponding drop off in the luminosity function for that obscuration model.

Our 1500Å predictions are shown in Figure 11 compared to observations from SDSS DR3 (Richards et al. 2006), 2SLAQ+SDSS (Croom et al. 2009), CFHQS+SDSS (Willott et al. 2010), NDWFS+DLS (Glikman et al. 2011), the COSMOS field (Ikeda et al. 2011; Masters et al. 2012), Subaru (Kashikawa et al. 2015) and SDSS Stripe 82 (Jiang et al. 2016). These have been converted to 1500Å - the conversions made are detailed in Appendix B. There is a strong
Figure 10. The soft X-ray luminosity function compared to observations. The dashed black line is the prediction without accounting for absorption effects, the solid black line is the prediction with the Hasinger (2008) visible fraction, the dotted black line is with the Aird et al. (2015) visible fraction and the blue line is with the observationally determined LZMH visible fraction. The observations are Hasinger et al. (2005) (circles) and Aird et al. (2015) (triangles).

Figure 11. The 1500Å luminosity function compared to observations which have been converted to 1500Å. The dashed black line is the prediction without accounting for absorption effects, the solid black line is the prediction with the Hasinger (2008) visible fraction, the dotted black line is with the Aird et al. (2015) visible fraction and the blue line is with my observationally determined LZMH visible fraction. The observations are Richards et al. (2006) (yellow triangles), Croom et al. (2009) (yellow circles), Willott et al. (2010) (red squares), Glikman et al. (2011) (blue circles), Ikeda et al. (2011) (red circles), Masters et al. (2012) (purple squares), Kashikawa et al. (2015) (red triangles) and Jiang et al. (2016) (blue squares).
dependence on the assumed obscuration model. Our predictions are a good fit at $z \approx 2$ if we adopt the Hasinger (2008) visible fraction, whereas our observationally determined LZMH model fits best for $z \approx 4$. The reason for this difference is likely to be because Hasinger (2008) fitted their obscuration model at lower redshift whereas we are trying to fit for $z = 0 - 6$ with our LZMH visible fraction. Thus, unsurprisingly, the different visible fractions are likely to fit better in different redshift ranges.

While the model gives an acceptable fit to observations of the optical luminosity function at $z = 4$, it overpredicts the number of objects compared to the observed luminosity function at $z = 6$. This is a result of the model not strongly evolving in the redshift interval $z = 4 - 6$, while the observations indicate a stronger evolution in this redshift interval (Jiang et al. 2016). This discrepancy could be due to a variety of reasons. We suggest two possible explanations for this discrepancy and two corresponding variants on the model which provide a better fit to the observations at $z = 6$.

Firstly, the discrepancy could be due to obscuration. At $z = 6$ the visible fraction is not constrained by any observations, and so in Figure 12 we present predictions with a lower visible fraction at $z = 6$, which give a better fit to the $z = 6$ optical luminosity function. We show predictions for the standard model with two obscuration models: the LZMH visible fraction and the Z6MH visible fraction (c.f. Section 2.9). The Z6MH visible fraction needed to fit $z = 6$ is about a quarter of the LZMH visible fraction at $z < 6$. Thus $z > 6$ QSOs could be much more obscured than $z < 6$ QSOs.

Secondly, the discrepancy could be due to black hole accretion being less efficient at high redshift. At high redshift, galaxies have smaller host dark matter haloes and therefore sit in shallower gravitational potential wells. This means that gas is more easily blown out from galaxies and so less gas could be available for black hole accretion, leading to a less efficient black hole accretion. We therefore present a model with parameters that have been modified compared to the original calibration on observed data at low redshift. We change the parameter $f_{BH}$, which sets the fraction of mass accreted onto a black hole in a starburst event and the parameter $\eta_{Edd}$, which controls the degree of super-Eddington luminosity suppression. In the fiducial model, $f_{BH} = 0.005$ and $\eta_{Edd} = 4$. $f_{BH} = 0.002$ and $\eta_{Edd} = 16$ give a better fit to the observations of the 1500Å luminosity function at $z = 6$ in Figure 12. However, we note that $\eta_{Edd} = 16$ means that there is very little super-Eddington luminosity suppression, whereas the ‘slim disc’ model for super-Eddington sources predicts significant super-Eddington luminosity suppression. We refer to this model as the ‘low accretion efficiency model’. In this model we use the LZMH visible fraction.

Figure 12. The 1500Å luminosity function at $z = 6$. We show predictions without obscuration (dashed black), with the Aird et al. (2015) visible fraction (dot-dash), with the ‘low z modified Hopkins’ visible fraction with the standard model (black solid), with the ‘$z = 6$ modified Hopkins’ visible fraction (dotted) and with the ‘low z modified Hopkins’ visible fraction with the different parameters (blue solid). The observations are from Willott et al. (2010) (red squares), Kashikawa et al. (2015) (red triangles) and Jiang et al. (2016) (blue squares).
5 EVOLUTION OF AGN LUMINOSITY FUNCTION AT Z > 6

In the left panel of Figure 13 the evolution of the AGN bolometric luminosity function for the standard model is shown for $7 < z < 15$. As the redshift increases, both the number of objects and the luminosities follow a downward trend. By $z \approx 12$, there are almost no objects brighter than $L_{\text{bol}} \sim 10^{46}\text{ergs}^{-1}$ in our simulated volume of $(800\text{Mpc})^3$. In the middle panel of Figure 13 we split the AGN luminosity function at $z = 9$ into the contribution from ADAFs, thin discs and super-Eddington objects. This shows how the trends that we saw in Figure 6 continue. As we go out to higher redshift, the contribution from ADAFs decreases, such that by $z = 9$, their contribution is very small. At low luminosities, the thin disc contribution just dominates over super-Eddington objects, while at high luminosities super-Eddington objects dominate. This suggests that most of the QSOs that will be detected by future surveys will be accreting above the Eddington accretion rate.

In the right panel of Figure 13 we split the AGN luminosity function at $z = 9$ by fuelling mode into hot halo mode, and starbursts triggered by mergers and disc instabilities. The dominant contributor at all luminosities at $z = 9$ is starbursts triggered by disc instabilities (in the figure the line for this is buried by the line for the total luminosity function), so we expect future surveys to detect AGN fuelled by this mechanism.

Surveys in the next decade will greatly enhance our understanding of the high redshift Universe. JWST will have instruments for both imaging and spectroscopy, including the NIRCam for optical to near-infrared imaging and MIRI for mid-infrared imaging. In parallel and over a period of six years, EUCLID will conduct two surveys: a ‘Wide Survey’ covering 15000 deg$^2$ of sky and a ‘Deep Survey’ covering 40 deg$^2$ from three fields. This will be conducted in four bands - one visible (VIS) and three near-IR (Y,J,H). In addition, ATHENA will detect X-rays using two instruments: the X-ray Integral Field Unit (X-IFU) for high resolution spectroscopy and the Wide Field Imager (WFI) for a large field of view with excellent survey power (Nandra et al. 2013).

We show predictions for JWST, EUCLID and ATHENA with three different models. First we have the standard model which uses the LHMF visible fraction (black solid line in Figure 12), one in which we vary the obscuration which uses the Z6MH visible fraction (black dotted line in Figure 12), and one low accretion efficiency model (blue solid line in Figure 12). The difference between the LHMF and Z6MH models is mostly in the number of objects detected, while there is a difference in the luminosities between the LHMF and the low accretion efficiency model. The low accretion efficiency model has only slightly fewer objects at lower luminosity than the LHMF model but has much fewer objects at higher luminosity compared to the LHMF model.

There are two effects that limit detections of $z > 7$ AGNs: flux and survey area limitations. The former affects the ability to detect low luminosity sources and the latter affects the number density down to which one can probe. From the predicted flux limits of the surveys, luminosity limits can be derived using $L = 4\pi d_l^2 f_r$ for calculating broadband luminosities (ATHENA) and $L_v = 4\pi d_l^2 f_r/(1 + z)$ for calculating a luminosity per unit frequency (EUCLID and JWST). $f_r$ is the flux per unit frequency and $d_l$ is the luminosity distance to the source, while $L$ is the luminosity in the rest-frame band/wavelength corresponding to the observed band/wavelength. We use these definitions to calculate luminosity limits (vertical lines) in Figures 14 to 19.

The number density limit of the survey can be calculated via the following method. The number of objects per log flux per solid angle per redshift is given by:

$$\frac{d^3N}{d\log L_v d\Omega dz} = \frac{d^2N}{d\log L_v dV} \frac{dV}{d\Omega dz}.$$  (48)

where $d^2N/d\log L_v dV$ is the luminosity function and $dV/d\Omega dz$ is the comoving volume per unit solid angle per unit redshift. For at least one detectable object per log flux, per unit redshift, we therefore have the condition:

$$\frac{d^2N}{d\log L_v dV} \geq \frac{1}{\pi d_l^2 \Delta \Omega}.$$  (49)

This condition allows us to construct the number density limits (horizontal lines) in Figures 14 to 19. Note that this limit is almost independent of redshift over the range $7 < z < 15$, as also seen for the JWST predictions of Cowley et al. (2018) for galaxies. The survey values adopted in our calculation are given in Table 4.

Luminosities shown have been K-corrected to a fixed band in the observer frame. Our template SED for this calculation is that of Marconi et al. (2004). To calculate the luminosity in each band we input the bolometric luminosity and the redshift and then integrate the SED over frequency with the appropriate response function for the filter in the rest frame of the source.

We show the predictions for three different JWST NIRCam bands. Note we do not make predictions for JWST MIRI, because we do not have a torus model to be able to model emission in the mid-infrared. Figure 14 is for the F070W (0.7μm) band, Figure 15 is for the F200W (2μm) band and Figure 16 is for the F444W (4μm) band. The sensitivities vary between the different bands with the F200W band having the best sensitivity of the three. Across the three different obscuration models, in the F200W band about one object per field of view at $z = 7$ should be detectable, which means that this band offers a good opportunity for detecting AGNs, which can then be compared to AGNs detected in X-rays to gain a multi-wavelength picture of distant AGNs. All the optical/near-IR JWST bands should be able to detect AGNs with a few fields of view out to $z = 10$, beyond which detection with JWST will become more difficult, as AGNs become rarer. Given that the interest is in viewing $z > 6$ quasars where the key SMBH growth is occurring, JWST should be able to probe these regions, although JWST offers less of an opportunity for detecting AGNs because of the small field of view and because AGNs are relatively sparse. For every 1000 fields of view imaged by JWST, we predict that 20 – 80 AGNs at $z = 7$ will be observed in the F070W band, 90 – 400 AGNs in the F200W band and 60 – 200 in the F444W band.

We show predictions for the EUCLID VIS (0.55-0.9μm) band and EUCLID H (1.5-2μm) band in Figures 17 and 18 respectively. We show the sensitivity and survey volume limits of the EUCLID Deep and EUCLID Wide surveys.
Figure 13. The predicted AGN bolometric luminosity function for the standard model at high redshift. Left panel: The evolution of the bolometric luminosity function for $z = 7$ (black), $z = 8$ (red), $z = 9$ (yellow), $z = 10$ (green), $z = 12$ (light blue), $z = 15$ (purple). The turnover at low luminosity is due to the halo mass resolution. Middle panel: The total AGN bolometric luminosity function at $z = 9$ (black) split into ADAFs (green), thin discs (purple) and super-Eddington objects (grey). Right panel: The total AGN bolometric luminosity function split into fuelling by hot halo mode (red), starbursts triggered by mergers (light blue) and starbursts triggered by disc instabilities (dark blue).

Table 4. The sensitivities and solid angles covered by the planned surveys. From https://jwst-docs.stsci.edu/display/JTI/NIRCam+ Sensitivity (JWST), https://www.euclid-ec.org/?page_id=2581 (EUCLID) and https://www.cosmos.esa.int/documents/400752/507693/Athena_SciRd_iss1v5.pdf (ATHENA)

| Instrument | Filter | $\lambda (\mu \text{m})$ or E(keV) | Sensitivity | Field of View | Assumed integration Time (ks) |
|------------|--------|-----------------------------------|-------------|---------------|-----------------------------|
| JWST NIRCam | F070W | 0.6 – 0.8 $\mu$m | 22.5 nJy | $2 \times 2.2 \times 2.2$ arcmin$^2$ | 10 |
| | F200W | 1.7 – 2.3 $\mu$m | 9.1 nJy | $2 \times 2.2 \times 2.2$ arcmin$^2$ | 10 |
| | F444W | 3.8 – 5.1 $\mu$m | 23.6 nJy | $2 \times 2.2 \times 2.2$ arcmin$^2$ | 10 |
| EUCLID (Deep) | VIS | 0.55 – 0.9 $\mu$m | 91.2 nJy | 40 deg$^2$ (survey) | - |
| | H | 1.5 – 2 $\mu$m | 145 nJy | 40 deg$^2$ (survey) | - |
| EUCLID (Wide) | VIS | 0.55 – 0.9 $\mu$m | 575 nJy | 150000 deg$^2$ (survey) | - |
| | H | 1.5 – 2 $\mu$m | 912 nJy | 150000 deg$^2$ (survey) | - |
| ATHENA WFI | Soft X-ray | 0.5 – 1.2 keV | $2.4 \times 10^{-17}$ erg cm$^{-2}$ s$^{-1}$ | 40 x 40 arcmin$^2$ | 450 |
| | Hard X-ray | 2 – 10 keV | $2.1 \times 10^{-16}$ erg cm$^{-2}$ s$^{-1}$ | 40 x 40 arcmin$^2$ | 450 |

Slightly more objects should be detected in the EUCLID near-IR bands than the EUCLID visible band: for the EUCLID Deep survey, in the VIS band at $z = 7$, 90 – 300 objects should be detected (depending on obscuration model, varying from the Z6MH to the LzMH visible fraction) and in the H band at $z = 7$, 100 – 500 objects should be detected. This is because the number density in the EUCLID near-IR band is greater for a given luminosity compared to the EUCLID visible band, despite EUCLID having better sensitivity in the visible band. For the EUCLID Wide survey, for the VIS band at $z = 7$, 5000 – 20000 objects should be detected, while in the H band at $z = 7$, 8000 – 30000 objects should be detected.

EUCLID near-IR bands at high redshift ($z = 10$) are closer to the peak of our AGN SED and so correspond to higher AGN luminosity compared to the visible band. At lower redshift ($z = 7$) the visible and near-IR bands give similar luminosities. Therefore assuming the template SED we have used, the near-IR bands will detect more $z = 10$ objects. Alternatively such observations may reveal that the AGN SED shape at high redshift is different to the AGN SED used in this work.

It will be difficult to detect very high redshift ($z = 15$) objects with EUCLID, such investigation may have to wait until surveys after EUCLID. This is because despite the survey area being sufficient for the number density of objects at this redshift, the sensitivity is not sufficient to detect these low luminosity objects. However, given the low number density of objects predicted at $z = 15$, which is a result of there being very few dark matter haloes and hence galaxies at this redshift, this redshift may not offer significant insights into the role of SMBHs in galaxy formation.

In Figure 19 we show the predictions for the ATHENA WFI in the soft X-ray (0.5-2 keV) band. We do not include obscuration for these soft X-ray predictions because at the redshifts we are considering, the rest frame band lies in the hard X-ray - a band for which we are assuming no obscura-
High redshift AGN predictions

Figure 14. Predictions for the AGN luminosity function in the JWST NIRCam F070W (0.7 µm) band. We show the fiducial model total luminosity function without obscuration (black dashed), the fiducial model with the ‘low z modified Hopkins’ visible fraction (black solid), the fiducial model with the ‘z = 6 modified Hopkins’ visible fraction (black dotted), and the low accretion efficiency model which uses the ‘low z modified Hopkins’ visible fraction (blue solid). The vertical lines are the luminosity limit caused by the sensitivity and the horizontal lines are the number density limit caused by the survey area. Detectable objects are therefore above and to the right of these lines. The upper horizontal line for each plot is for one object per field of view and the lower horizontal line is for one object per $10^3$ fields of view.

Figure 19 also shows the predictions for ATHENA in the hard X-ray (2-10 keV) band (blue lines). In our template SED, an AGN emits more at hard X-ray energies than at soft X-ray energies, but the minimum luminosity that can be detected is higher for the hard X-ray band because the hard X-ray effective area of ATHENA is smaller than at the soft X-ray band by approximately an order of magnitude. This means that the number of objects expected to be detected in the hard X-ray band is not as many as the soft X-ray band. We predict 5 – 10 QSOs will be detected at $z = 7$ in the hard X-ray band per field of view, whereas we predict 30 – 70 QSOs will be detected at $z = 7$ in the soft X-ray band per field of view, depending on the model. The low accretion efficiency model predicts fewer objects across all luminosities and redshifts.

The alternative models featuring a lower visible fraction or low accretion efficiency predict fewer objects, so observations using ATHENA, EUCLID and JWST should be able to differentiate between these models and thus provide better understanding of the high redshift population.

We show the SMBH masses, dimensionless mass accretion rates and host galaxy stellar masses of the objects detectable with each instrument at redshifts $z = 7$ and 10. in Table 5. Compared to EUCLID, JWST will probe smaller SMBHs in smaller host galaxies with lower dimensionless mass accretion rates. The two different EUCLID surveys will detect different populations of objects, with EUCLID Wide detecting larger SMBHs, in larger host galaxies, with higher dimensionless mass accretion rates than EUCLID Deep. ATHENA also offers the detection of a slightly different set of objects with the typical host galaxy masses, black hole masses and dimensionless accretion rates similar to the values for EUCLID Deep at $z = 7$ but with lower values of these quantities compared to EUCLID Deep at $z = 10$. For each survey, the objects detected at $z = 10$ have lower black hole masses, lower stellar masses and higher dimensionless accretion rates than at $z = 7$. Comparing all the telescopes at $z = 7$, the JWST F200W band will detect the smallest mass black holes while the ATHENA hard X-ray band will detect the largest mass black holes. The VIS band for the EUCLID Wide survey will detect the objects with the highest dimensionless mass accretion rates (therefore giving us insights into the population of super-Eddington objects), while the JWST F200W band will detect the objects with the lowest dimensionless mass accretion rates. The VIS band for the EUCLID Wide survey band will detect objects in the highest mass host galaxies, while the JWST F200W band will detect objects with the lowest mass host galaxies.
Figure 15. As above but for the JWST NIRCam F200W (2μm) band.

Figure 16. As above but for the JWST NIRCam F444W (4.44μm) band.
Figure 17. Predictions for the AGN luminosity function in the EUCLID VIS (550-900 nm) observer frame band. The long dashed lines represent the sensitivity and survey volume limits of the EUCLID Deep survey and the short dashed lines represent the sensitivity and survey volume limits of the EUCLID Wide survey.

Figure 18. The same as Figure 17 but for the EUCLID H (1.5-2\mu m) band.
of the objects detected by JWST, EUCLID and ATHENA. The ATHENA hard X-ray band will not be able to detect AGN at $z$.

Table 5. The medians, 16th and 84th percentiles of the SMBH masses, dimensionless mass accretion rates and host galaxy stellar masses of the objects detected by JWST, EUCLID and ATHENA. The ATHENA hard X-ray band will not be able to detect AGN at $z = 10$.

| Redshift | Instrument | Filter | $M_{\text{SMBH}}(M_\odot)$ | $\dot{m} = M/M_{\text{Edd}}$ | $M_*(M_\odot)$ |
|----------|------------|--------|----------------|-----------------|--------------|
| $z = 7$  | JWST       | F070W  | $8.3^{+14.8}_{-3.9} \times 10^6$ | 0.9$^{+1.1}_{-0.3}$ | 1.8$^{+2.1}_{-0.8} \times 10^9$ |
|          |            | F200W  | $2.1^{+1.2}_{-1.1} \times 10^6$ | 0.7$^{+0.4}_{-0.2}$ | 5.9$^{+2.7}_{-1.6} \times 10^8$ |
|          |            | F444W  | $3.2^{+2.7}_{-1.6} \times 10^6$ | 0.8$^{+0.4}_{-0.3}$ | 8.4$^{+3.8}_{-3.8} \times 10^8$ |
|          | EUCLID Deep| VIS    | $2.0^{+2.5}_{-0.9} \times 10^7$ | 1.6$^{+2.7}_{-0.7}$ | 3.5$^{+4.4}_{-1.6} \times 10^9$ |
|          |            | H      | $1.6^{+2.5}_{-0.7} \times 10^7$ | 1.4$^{+2.3}_{-0.6}$ | 3.0$^{+4.2}_{-1.4} \times 10^9$ |
|          | EUCLID Wide| VIS    | $4.3^{+5.5}_{-3.3} \times 10^7$ | 4.0$^{+5.3}_{-2.0}$ | 5.9$^{+4.9}_{-2.7} \times 10^9$ |
|          |            | H      | $3.8^{+4.8}_{-2.7} \times 10^7$ | 3.3$^{+3.9}_{-1.6}$ | 5.4$^{+4.9}_{-2.7} \times 10^9$ |
|          | ATHENA WFI | Soft X-ray | $9.2^{+15.9}_{-5.2} \times 10^6$ | 1.0$^{+1.1}_{-0.3}$ | 1.9$^{+2.3}_{-0.8} \times 10^9$ |
|          |            | Hard X-ray | $2.9^{+3.0}_{-1.3} \times 10^7$ | 2.4$^{+1.2}_{-1.2}$ | 4.4$^{+2.1}_{-1.2} \times 10^9$ |
| $z = 10$ | JWST       | F070W  | $5.1^{+2.8}_{-1.4} \times 10^6$ | 2.3$^{+1.5}_{-0.9}$ | 7.9$^{+4.6}_{-2.3} \times 10^8$ |
|          |            | F200W  | $1.7^{+0.7}_{-0.3} \times 10^6$ | 1.3$^{+0.4}_{-0.4}$ | 3.5$^{+1.6}_{-1.6} \times 10^8$ |
|          |            | F444W  | $2.4^{+1.8}_{-1.0} \times 10^6$ | 1.4$^{+0.9}_{-0.9}$ | 4.5$^{+2.1}_{-2.1} \times 10^8$ |
|          | EUCLID Deep| VIS    | $1.1^{+0.8}_{-0.3} \times 10^7$ | 5.3$^{+3.1}_{-2.1}$ | 1.2$^{+1.0}_{-0.6} \times 10^9$ |
|          |            | H      | $8.9^{+4.3}_{-2.4} \times 10^6$ | 4.4$^{+2.0}_{-1.9}$ | 1.0$^{+0.9}_{-0.5} \times 10^9$ |
|          | EUCLID Wide| VIS    | $2.6^{+1.5}_{-1.2} \times 10^7$ | 13.6$^{+2.0}_{-2.0}$ | 2.0$^{+1.4}_{-1.4} \times 10^9$ |
|          |            | H      | $1.9^{+1.4}_{-0.8} \times 10^7$ | 10.5$^{+1.8}_{-5.4}$ | 1.7$^{+1.4}_{-0.8} \times 10^9$ |
|          | ATHENA WFI | Soft X-ray | $4.9^{+3.3}_{-2.0} \times 10^6$ | 2.3$^{+1.7}_{-1.0}$ | 7.6$^{+3.0}_{-4.0} \times 10^8$ |
|          |            | Hard X-ray | -            | -                | -              |
Understanding the evolution of AGN across cosmic time has been of interest ever since they were discovered to reside outside of our own galaxy. AGN have also been shown to be important in how galaxies evolve - particularly given the dramatic effect they have on their host galaxies through AGN feedback. However, many uncertainties remain, such as the nature of the physical processes involved in AGN feedback and SMBH spin evolution. The origin of SMBHs and their role in the distant Universe still remains a mystery.

Fortunately the next decade offers us an exciting new opportunity to probe the high redshift Universe, especially given the plans for powerful new telescopes such as JWST, EUCLID and ATHENA. They will offer us a multiwavelength view of the distant Universe (visible, near-IR, X-ray) and allow us to characterise poorly understood physical processes. The role of SMBHs and their growth in the distant Universe will be probed with much greater accuracy than before.

With these potential new developments in mind, we present predictions for the high redshift Universe using a high volume, high resolution dark matter simulation (P-Millennium) populated with galaxies using the semi-analytic model of galaxy formation GAlform. This updated scheme for the SMBH spin evolution is used within the Lacey et al. (2016) GAlform model, which has been shown to reproduce a large number of observable galaxy properties over an unprecedented wavelength and redshift range. The model that we use incorporates an updated prescription for SMBH spin evolution: for these predictions we have assumed SMBH spin evolving in a ‘chaotic accretion’ scenario in which the angle between the accretion disc and the SMBH spin randomises once a self-gravity mass of gas has been consumed.

We then calculated AGN bolometric luminosities from the SMBH mass accretion rate, taking into account the SMBH spin and the different radiative efficiencies for different accretion regimes (ADAFs, thin discs, super-Eddington objects). Then using a template SED and different obscuration models we derived AGN luminosities in the hard X-ray, soft X-ray and optical/UV (1500 Å) bands.

The model is consistent with both the observed black hole mass functions and SMBH versus bulge mass correlations. We compare the AGN luminosity functions in the redshift range $0 < z < 6$ to a wide range of observations at different wavelengths. The model is in good agreement with the observations. We split the luminosity functions by accretion mode (ADAFs, thin discs, super-Eddington objects) and by fuelling mode (hot halo or starbursts triggered by disk instabilities or mergers) to see the relative contributions. At low redshifts and low luminosities the ADAF contribution dominates but at higher luminosities and higher redshifts, the thin disc and super-Eddington objects dominate the luminosity function. Hot halo mode fuelled accretion dominates at low redshift and low luminosity, but at higher redshift and higher luminosity, starbursts triggered by disc instabilities dominate the luminosity function.

We study the bolometric luminosity function at high redshift ($7 < z < 15$) and split by accretion mode and fuelling mode - the dominant accretion modes are thin discs at low luminosities, and super-Eddington objects at high luminosities, and the dominant fuelling mode is starbursts triggered by disc instabilities.

We then present predictions for JWST, EUCLID and ATHENA and show these alongside sensitivity and survey volume limits. We also present two alternative predictions that provide a better fit to the $z = 6$ optical luminosity function - these predictions separetely vary the amount of AGN obscuration and the SMBH accretion efficiency, which are uncertainties for the AGN population at high redshift. Comparing these predictions to observations should allow us to be able to better constrain obscuration and SMBH accretion efficiency at high redshift.

We find that the three different surveys will detect different samples of objects. The objects detected at $z = 7$ will have median black hole masses that vary from $2 \times 10^{6} M_{\odot}$ to $4 \times 10^{7} M_{\odot}$, median dimensionless mass accretion rates that vary from $1 - 4$ and median host galaxy stellar masses that vary from $6 \times 10^{8} M_{\odot}$ to $6 \times 10^{9} M_{\odot}$, providing different but complementary views on the $z > 6$ AGN population. Many super-Eddington objects should be detectable and SMBHs not much bigger than the largest SMBH seed mass should be detected.

There are many natural continuations from this work. We have already mentioned that we only assume a quasar SED for our bolometric corrections, in reality we have a variety of AGNs having different accretion rates in different accretion regimes, which will have different SEDs (e.g. Jin et al. 2012). Combining different template SEDs for different regimes may allow the model to predict luminosity functions in greater agreement with the observations. Secondly, we may wish to more thoroughly explore the dependence of the model on the SMBH spin evolution model used e.g. investigating the difference between accreting in increments of the self-gravity mass and increments of the warp mass. Finally, in this paper we do not show radio luminosity functions - given that AGN jets are observed to have a strong effect on their host galaxies and given that these jets emit at radio wavelengths via synchrotron emission, an investigation into radio emission would also be important for understanding the role of AGN in galaxy formation.

ACKNOWLEDGEMENTS

This work was supported by the Science and Technology facilities Council ST/L00075X/1. AJG acknowledges an STFC studentship funded by STFC grant ST/N50404X/1. CL is funded by an Australian Research Council Discovery Early Career Researcher Award (DE150100618) and by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K009422/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.
APPENDIX A: EFFECTS OF VARYING SMBH SEED MASS

In Figure A1 we show the effect of varying the SMBH seed mass on the black hole mass function at $z = 0$. We show plots for SMBH seed masses of $10^{-1}M_\odot$ (the default value), $10^3h^{-1}M_\odot$ and $10^5h^{-1}M_\odot$. Generally the black hole mass function reaches a converged value at about 100 times the black hole seed mass. We also plot as vertical lines $M_{\text{BH}} = M_{\text{seed}}$, $M_{\text{BH}} = 2 \times M_{\text{seed}}$ and $M_{\text{BH}} = 3 \times M_{\text{seed}}$. It can be seen that the spikes in the black hole mass function occur at these values due to SMBH seeds merging with other SMBH seeds.

This convergence in properties at around 100 times the seed mass can also be seen in Figure A2, where the median of the SMBH mass vs bulge mass relation for seeds of mass $10^5h^{-1}M_\odot$ only converges with that for the other seed masses for SMBH masses above around $10^7M_\odot$. In Figure A3 we show the AGN bolometric luminosity function at $z = 12$ for the same three different seed masses. The luminosity functions for the three different seed masses are consistent with each other within statistical errors for $L_{\text{bol}} > 10^{43}$ ergs$^{-1}$.
APPENDIX B: CALCULATING BROAD-BAND OPTICAL MAGNITUDES FOR AGN

We define the filter-averaged luminosity per unit frequency $L_\nu$ for a filter R, redshifted into the rest-frame of the object at redshift $z$ as:

$$<L_\nu>_{R}^{(z)} = \frac{\int L_\nu (1+z) \nu_o R(\nu_o) d\nu_o}{\int R(\nu_o) d\nu_o},$$  \hspace{1cm} (B1)$$

where $R(\nu_o)$ is the response function of the filter at observed frequency $\nu_o$. The absolute magnitude in the observer frame band defined by the filter R is then defined as:

$$M_i^{AB,R} = -2.5 \log_{10} \left( \frac{<L_\nu>_{R}^{(z)}}{L_\nu^{o}} \right).$$  \hspace{1cm} (B2)$$

where $L_\nu^{o} = 4\pi (10pc^2) \times f_\nu^{o}$, with $f_\nu^{o} = 3631$Jy, the flux corresponding to an apparent AB magnitude of 0, and $L_\nu^{o}$ the corresponding luminosity for an absolute AB magnitude of 0. We remind readers that monochromatic AB (Absolute Bolometric) apparent magnitudes are defined using the following relation (Oke & Gunn 1983):

$$m_{AB}(\nu) = -2.5 \log_{10} \left( \frac{f_\nu}{3631 Jy} \right),$$  \hspace{1cm} (B3)$$

where $f_\nu$ is the observed flux of the source, which is related to the luminosity in the rest-frame of the object as:

$$f_\nu(\nu_o) = \frac{(1+z)L_\nu(1+z)\nu_o}{4\pi d_L^2}.$$  \hspace{1cm} (B4)$$

The apparent and absolute magnitudes are then related by

$$m_{AB}(\nu) = M_{AB}(1+z) - 2.5 \log_{10}(1+z) + 5 \log_{10}(d_L / 10pc).$$  \hspace{1cm} (B5)$$

We then use the following formulae to convert magnitudes in the observational data from different wavelengths given to rest-frame wavelength 1500Å. Note that we do not make a correction for emission lines, as done in Richards et al. (2006) - we are only comparing continuum luminosities in this study, which is consistent with the Marconi et al. (2004) continuum template used throughout this paper. The observations in Richards et al. (2006) are in the K-corrected SDSS i-band at $z = 2$, $M_i(z = 2) = M_i(z = 2) - 2.5 \log(1+z)$, where $M_i(z = 2)$ is the absolute magnitude at the rest-frame wavelength corresponding to the observed i-band at $z = 2$ as in equation (B2). To convert from $M_i(z = 2)$ to 1500Å we follow Richards et al. (2006) but using the spectral index value (defined by $L_\nu \propto \nu^{\alpha_o}$) of $\alpha_o = -0.44$ from Marconi et al. (2004) instead of $\alpha_o = -0.5$ in Richards et al. (2006). First we convert from $M'_i(z = 2)$ to $M_i(z = 0)$:

$$M_i(z = 0) = M_i'(z = 2) + 2.5(1 + \alpha_o) \log(1 + 2)$$  \hspace{1cm} (B6)$$

where $M_i(z = 0)$ is the absolute magnitude at the central wavelength of the i-band (7471Å). Then we convert from $M_i(z = 0)$ to the absolute magnitude at rest-frame 1500Å, $M_{1500}$, to give the correction to $M'_i(z = 2)$:

$$M_{1500} = M_i(z = 0) + 2.5\alpha_o \log_{10} \left( \frac{1500Å}{7471Å} \right),$$  \hspace{1cm} (B8)$$

$$M_{1500} = M_i(z = 0) + 0.767,$$  \hspace{1cm} (B9)$$

$$M_{1500} = M'_i(z = 2) + 1.435.$$  \hspace{1cm} (B10)$$

Jiang et al. (2009); Willott et al. (2010); Ikeda et al. (2011); Masters et al. (2012); Kashikawa et al. (2015) report observations at 1450Å, and define their absolute magnitudes without the extra redshift factor included in the Richards et al. (2006) definition. These absolute magnitudes at 1450Å, $M_{1450}$, can be converted to 1500Å using:

$$M_{1500} = M_{1450} + 2.5\alpha_o \log_{10} \left( \frac{1500Å}{1450Å} \right),$$  \hspace{1cm} (B11)$$

$$M_{1500} = M_{1450} - 0.016.$$  \hspace{1cm} (B12)$$

Finally Croom et al. (2009) report observations in the SDSS g-band (4670Å) K-corrected to $z = 2$ so we use the correction in their paper:

$$M'_g(z = 2) = M'_i(z = 2) + 2.5\alpha_o \log_{10} \left( \frac{4670Å}{7471Å} \right),$$  \hspace{1cm} (B13)$$

and combine it with the above relation giving:

$$M_{1500} = M'_g(z = 2) + 1.211.$$  \hspace{1cm} (B14)$$

APPENDIX C: VISIBLE AND OBSCURED FRACTIONS FOR AGN

The AGN visible fractions (the fraction of sources at a particular luminosity and redshift that are visible) derived in this paper have been obtained by constructing an observational bolometric luminosity function from observed luminosity functions at X-ray and optical wavelengths. These luminosities were converted to bolometric using the Marconi et al. (2004) AGN SED, and then the number densities were converted to bolometric using visible fractions of a functional form similar to Hopkins et al. (2007) dependent only on $L_{bol}$ (c.f. equation (41)). It was assumed that there is no obscuration for hard X-ray wavelengths. The coefficients of the visible fractions were then selected (c.f. equations (44), (45) and (46)) such that the bolometric luminosity function has the smallest scatter.

To construct a bolometric luminosity function from multiple sets of observations in different wavebands, different authors use different template SEDs. Some authors include reprocessed radiation from dust (its inclusion causes an ‘IR bump’ in the SED) whereas some do not. Including reprocessed radiation gives observed bolometric luminosities, whereas not including the IR bump gives intrinsic bolometric luminosities. The intrinsic bolometric luminosities are isotropic, while the observed bolometric luminosities are...
Figure C1. Comparing the visible fractions of different obscuration models. Shown are Hopkins et al. (2007) (black), Hasinger (2008) (light blue), Aird et al. (2015) (red), the LZMH model (dark blue) and the Z6MH model (purple). The solid lines for the observational visible fractions indicate the regimes where there is observational data, while the dotted lines indicate regions where a functional form has been extrapolated.

Figure C2. The same as the previous plot, but for soft X-ray.
Figure C3. The bolometric luminosity function derived in this work (blue) by using the Marconi et al. (2004) bolometric corrections, and by varying the coefficients of the visible fractions to obtain a bolometric luminosity function with the smallest scatter between points derived from data at different wavelengths, compared to the Hopkins et al. (2007) bolometric luminosity function (red). The Hopkins et al. (2007) bolometric luminosities have been multiplied by 7.9/11.8 to account for the different SED used (see text).

Figure C4. Comparing the effect of using different obscuration models on the constructed bolometric luminosity function. The left panels are using the obscuration model presented in this work, and the right panels are using the obscuration model for Hopkins et al. (2007). The upper panels are at $z = 0.2$ and the lower panels are at $z = 2$. 
not isotropic because the torus is not isotropic. Also, obscuration from the torus causes some AGNs not to be observed, which affects the number densities, so the number density of the observed bolometric luminosities will be different to the number density of the intrinsic bolometric luminosities, with the intrinsic bolometric luminosities reflecting the true number of objects. Therefore, we compare our predictions to estimates of the intrinsic bolometric luminosity function. Therefore the observed bolometric luminosity function in Hopkins et al. (2007) have had their luminosities multiplied by a factor 7/8 (Marconi et al. 2004) to account for this effect.

We show a comparison of the different optical obscuration models at 1500Å in Figure C1 and at soft X-ray energies in Figure C2. There is clearly still some uncertainty in the visible fraction.

The derived bolometric luminosity function is shown compared to the bolometric luminosity functions derived in Hopkins et al. (2007) in Figure C3. The bolometric luminosity function derived in this work is also similar to that determined by Shankar et al. (2009). These observationally determined visible fractions are redshift independent by construction. We explored whether a better fit could be obtained by including a redshift dependence. To obtain a better fit, the visible fraction needed to increase and then decrease (c.f. the redshift dependence derived by Aird et al. 2015), but even with a functional form to allow this, the scatter in the bolometric luminosity function was only slightly less than the redshift independent version.

To quantify the effect of the new obscuration model used for this paper we show the bolometric luminosity function derived using the Hopkins et al. (2007) visible fraction, compared to the bolometric luminosity function derived using the visible fraction presented in this paper in Figure C4. The new obscuration fraction does improve the constructed obscuration fraction, this reduction in scatter can be seen particularly at \( L_{\text{bol}} \sim 10^{44}\text{ergs}^{-1} \) at \( z = 0.2 \) and at \( L_{\text{bol}} \sim 10^{48}\text{ergs}^{-1} \) at \( z = 2 \).

**APPENDIX D: EXPLORING THE EFFECT OF VARYING PARAMETERS**

We show the effect on the bolometric luminosity function of varying some of the free parameters used in the model; in Figure D1, we show the effect of varying the parameter \( f_q \) (c.f. equation (26)). \( f_q \) affects the value of \( M \) and therefore the AGN luminosities. One expects a higher value of \( f_q \) to lead to lower values of \( M \) and therefore a steeper luminosity function at the bright end as we see in Figure D1. At the faint end, a lower value of \( f_q \) results in a poorer fit to the observations at low redshift (\( z = 0.2, 0.5, 1 \)) but is a better fit to the observations at high redshift (\( z = 2, 4, 6 \)). At the bright end, a higher value of \( f_q \) seems to give a better fit to the observations at low redshift but gives a worse fit to the observations at high redshift (e.g. around \( L_{\text{bol}} \sim 10^{44}\text{ergs}^{-1} \) at \( z = 4 \)). With these considerations in mind, we decide to keep the Fanidakis et al. (2012) value of \( f_q = 10 \) for our predictions in this paper.

We show the effect of varying the parameter \( \eta_{\text{ADAF}} \) (c.f. equation (32)) in Figure D2. \( \eta_{\text{ADAF}} \) controls the suppression of the luminosity for super-Eddington accretion rates, where a low value of \( \eta_{\text{ADAF}} \) corresponds to stronger luminosity suppression than a high value of \( \eta_{\text{ADAF}} \). This parameter only affects the very bright end of the luminosity function as we would expect. This parameter also has more of an effect at high redshift, where there are more super-Eddington sources. A value of \( \eta_{\text{ADAF}} = 1 \) gives a slightly better fit to the bright end observations at \( z = 6 \) but \( \eta_{\text{ADAF}} = 16 \) gives a better fit to bright end observations at \( z = 2 \) and \( z = 4 \). Therefore we once again to opt to keep the Fanidakis et al. (2012) value of \( \eta_{\text{ADAF}} = 4 \) for our predictions in this paper.

We show the effect of switching off disc instabilities in Figure D3. We show the fiducial model alongside a model in which all discs are stable and so no disc instability starbursts occur. Disc instabilities dominate the AGN luminosity function at \( z > 2 \), and so this is the regime where we expect turning off disc instabilities to have the most effect. For \( L_{\text{bol}} < 10^{46}\text{ergs}^{-1} \), at \( z > 2 \) switching off disc instabilities results in fewer starbursts and so there are fewer objects at these luminosities. For \( L_{\text{bol}} > 10^{46}\text{ergs}^{-1} \), at \( z > 2 \) the two models are similar - this is because if we switch off disc instabilities, mergers will trigger the starbursts that would have happened due to disc instabilities. At \( z < 2 \), switching off disc instabilities makes the luminosity function less steep.

We show the effect of switching off the accretion and merger spinup in Figure D4. The radiative accretion efficiency given to the black holes is that which corresponds to a black hole spin of \( a = 0.1 \). The luminosity functions for the two models are generally similar, although the fiducial model has a slightly lower number density at high luminosities.

We show the effect of changing the assumptions for accretion efficiency, \( \epsilon \), in Figure D5. We compare the fiducial model to a model in which the accretion efficiency is the thin disc accretion efficiency for all values of the specific mass accretion rate, \( \dot{m} \). Interestingly, this result provides a slightly better fit to the bolometric luminosity function, particularly for \( z < 0.5 \) and \( L_{\text{bol}} < 10^{46}\text{ergs}^{-1} \), where the fiducial model underpredicts the number density. This is the regime where ADAFs dominate the luminosity function, and so this test suggests that a better fit to the observed AGN luminosity function might be obtained if the radiative accretion efficiency for ADAFs is higher than the values assumed in our standard model.

This paper has been typeset from a \TeX/\LaTeX\ file prepared by the author.
Figure D1. Exploring the effect on the AGN bolometric luminosity function of varying the parameter $f_q$. Shown are $f_q = 5$ (blue), $f_q = 10$ (purple, the fiducial model) and $f_q = 20$ (red). The shading shows the Poisson errors of the distribution.

Figure D2. Exploring the effect of varying $\eta_{\text{Edd}}$. Shown are $\eta_{\text{Edd}} = 1$ (blue), $\eta_{\text{Edd}} = 4$ (purple, the fiducial model) and $\eta_{\text{Edd}} = 16$ (red).
Figure D3. Exploring the effect of switching off disc instabilities. Shown are the fiducial model (solid) and the model with disc instabilities switched off (dashed).

Figure D4. Exploring the effect of turning off the SMBH spinup evolution: the model with chaotic mode accretion spinup and merger spinup (red) and the model with no accretion nor merger spinup with a thin disc accretion efficiency, $\epsilon_{TD} = 0.1$ (blue).
Figure D5. Exploring the effect of changing the accretion efficiency $\epsilon$: the model with $\epsilon = \epsilon_{TD}$ as the accretion efficiency for all $\dot{m}$ regimes (blue) and the fiducial model (red).