

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society ©: 2019 The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Active galactic nuclei and their large-scale structure: an eROSITA mock catalogue

J. Comparat,¹★ A. Merloni,¹★ M. Salvato¹,¹★ K. Nandra,¹ T. Boller,¹
 A. Georgakakis², A. Finoguenov,³ T. Dwelly¹, J. Buchner^{4,5,6}, A. Del Moro,¹
 N. Clerc,⁷ Y. Wang,⁸ G. Zhao^{8,9,10}, F. Prada,¹¹ G. Yepes,¹² M. Brusa,^{13,14}
 M. Krumpe¹⁵ and T. Liu¹

¹Max-Planck-Institut für extraterrestrische Physik (MPE), Giessenbachstrasse 1, D-85748 Garching bei München, Germany

²Institute for Astronomy & Astrophysics, National Observatory of Athens, V. Paulou & I. Metaxa, 11532, Greece

³Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2a, FI-00014 Helsinki, Finland

⁴Pontificia Universidad Católica de Chile, Instituto de Astrofísica, Casilla 306, Santiago 22, Chile

⁵Excellence Cluster Universe, Boltzmannstr. 2, D-85748, Garching, Germany

⁶Millennium Institute of Astrophysics, Vicuña MacKenna 4860, 7820436 Macul, Santiago, Chile

⁷IRAP, Université de Toulouse, CNRS, UPS, CNES, Toulouse, 31400, France

⁸National Astronomy Observatories, Chinese Academy of Science, Beijing, 100101, P.R. China

⁹School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, P.R. China

¹⁰Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

¹¹Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain

¹²Departamento de Física Teórica and CIAFF, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

¹³Dipartimento di Fisica e Astronomia, Alma Mater Studiorum Università di Bologna, via Gobetti 93/2, I-40129 Bologna, Italy

¹⁴INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy

¹⁵Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany

Accepted 2019 May 15. Received 2019 May 14; in original form 2019 January 30

ABSTRACT

In the context of the upcoming SRG/eROSITA survey, we present an N -body simulation-based mock catalogue for X-ray-selected active galactic nucleus (AGN) samples. The model reproduces the observed hard X-ray AGN luminosity function (XLF) and the soft X-ray logN–logS from redshift 0 to 6. The XLF is reproduced to within ± 5 per cent and the logN–logS to within ± 20 per cent. We develop a joint X-ray – optical extinction and classification model. We adopt a set of empirical spectral energy distributions to predict observed magnitudes in the UV, optical, and NIR. With the latest eROSITA all sky survey sensitivity model, we create a high-fidelity full-sky mock catalogue of X-ray AGN. It predicts their distributions in right ascension, declination, redshift, and fluxes. Using empirical medium resolution optical spectral templates and an exposure time calculator, we find that 1.1×10^6 (4×10^5) fibre-hours are needed to follow-up spectroscopically from the ground the detected X-ray AGN with an optical magnitude $21 < r < 22.8$ ($22.8 < r < 25$) with a 4-m (8-m) class multiobject spectroscopic facility. We find that future clustering studies will measure the AGN bias to the per cent level at redshift $z < 1.2$ and should discriminate possible scenarios of galaxy-AGN co-evolution. We predict the accuracy to which the baryon acoustic oscillation standard ruler will be measured using X-ray AGN: better than 3 per cent for AGN between redshift 0.5 to 3 and better than 1 per cent using the Ly α forest of X-ray QSOs discovered between redshift 2 and 3. eROSITA will provide an outstanding set of targets for future galaxy evolution and cosmological studies.

Key words: galaxies: active – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

Within the next few years, the eROSITA X-ray telescope on board the Spectrum-Roentgen-Gamma (SRG) mission will detect and

* E-mail: comparat@mpe.mpg.de (JC); am@mpe.mpg.de (AM); mara@mpe.mpg.de (MS)

measure the position and properties of about 3 million active galactic nuclei (AGNs) and 100 000 clusters of galaxies over the full sky (Merloni et al. 2012; Predehl et al. 2016). Such large numbers of AGN will enable us to compute accurately at least their one- and two-point distribution functions over cosmological volumes, and thereby tightly interconnect the evolution of growing supermassive black holes in the overall galaxy population with the large-scale structure of dark matter (Fanidakis et al. 2011). For example, the upcoming clustering measurements of the most luminous AGN over the full sky should be precise enough to retrieve their corresponding halo population parameters to within better than 30 per cent (~ 0.1 dex on halo masses).

To enable these clustering measurements, large-scale cosmological simulations are a key tool for three main reasons: First, they enable a thorough validation of the method used to construct the catalogues and understand possible systematic errors. They constitute an input for end-to-end simulations of the catalogue creation process (e.g. Reid et al. 2016). Secondly, they allow the construction and constraint of the possible halo occupation distribution of AGN and represent a key tool to assess possible systematic errors on the measurements (e.g. Ross et al. 2017). Finally, they provide the necessary information to create covariance matrices with respect to cosmological parameters or halo occupation distribution parameters (e.g. Klypin & Prada 2018).

This paper is a continuation of Georgakakis et al. (2018, hereafter G18), where a model linking AGN to dark matter haloes was explained and tested against a set of clustering measurements of X-ray-selected AGN available in the literature. We present here a generalization of this method to redshift $0 < z < 6$ that we project on the full sky by including an accurate model of the eROSITA all-sky survey selection function. The mock catalogue presented here is a cornerstone to start the eROSITA end-to-end data processing validation tasks.

The structure of the paper is as follows. We discuss the cosmological N -body simulations used to create the mock in Section 2. Then, we describe the analytic model we created to simulate the AGN population in Section 3; we explain how AGN populate dark matter haloes and how their X-ray emission is calculated; we detail the dust extinction model used to predict both the X-ray and the optical emission of AGN. In Section 4, we present a set of forecasts generated using the mock catalogue itself.

First, we give the expected redshift distribution of eROSITA AGN as a function of survey depth and AGN type. Then, we lay out a possible spectroscopic follow-up strategy (see Section 4.2). In Section 4.3, we detail how well the X-ray luminosity function (XLF) will be recovered at the bright end, while in Section 4.4 we forecast the accuracy with which the baryon acoustic oscillation standard ruler will be measured throughout the redshift range $0.5 < z < 4$. In Section 4.5, we detail to what accuracy the large-scale halo bias of AGN will be measured. In Section 5, we detail a possible strategy towards completing eROSITA end-to-end simulations.

We assume a flat Λ CDM cosmology close to that of the Planck Collaboration et al. (2014) with $\Omega_m = 0.307115$, $h = 0.6777$, $\sigma_8 = 0.8228$. Magnitudes are in the AB system. Throughout the paper L_X refers to the X-ray luminosity in the 2–10 keV band. The set of eROSITA mocks is made public here.¹

2 N-BODY DATA

We use the halo catalogues of the MDPL2 simulation² (Planck cosmology, 3840³ particles of mass $2.2 \times 10^9 M_\odot$, 1475.5 Mpc on the side, Klypin et al. 2016) post-processed with the ROCKSTAR merger trees software (Behroozi, Wechsler & Wu 2013). The halo mass function is correct to a few percent down to halo masses of $M_{\text{vir}} \sim 10^{11} M_\odot$ (Comparat et al. 2017). In this analysis, we consider distinct haloes more massive than $2 \times 10^{11} M_\odot$. This resolution is sufficient to resolve haloes that contain the least massive AGN to be observed by eROSITA; indeed, G18 showed that a higher resolution simulation would not bring additional information on cosmological scales. Alternative simulations that could be used for this project include those of Angulo et al. (2012), Skillman et al. (2014), Heitmann et al. (2015), and Ishiyama et al. (2015).

The configurations used to cover the redshift range $0 < z < 6$ are given in Table 1. We replicate 6 times the box along the line of sight to reach up to redshift 6. Each replicated box is sliced (like an onion) into shells according to the availability of snapshots. There are about 20 snapshots per replica, implying an approximate width of about 80 Mpc per snapshot. This corresponds to a time-step between shells of 50 Myr at redshift 6 that increases to 250 Myr at redshift 0. The shells and replication scheme are illustrated in Fig. 1. We then pave these shells on the full sky using rotation matrices. We compute the new position and velocity vectors for each halo and its projection in right ascension, declination, and redshift (real and redshift space). We then populate or ‘paint’ the full-sky mock with AGN, as we describe in the following section.

3 AGN MODEL

We create an analytic model of the population of X-ray emitting supermassive black holes, which we subsequently refer to as ‘AGN’ [see Padovani et al. (2017) for extended discussions about the AGN naming conventions]. The aim of this model is to follow accurately, within 10–20 per cent, the observed hard XLF and the number density projected on the sky as a function of soft X-ray flux, while tracing the large-scale structure of the Universe. We use as starting point semi-analytical models (SAMS; e.g. Lacey et al. 2016), which reproduce quite accurately the bulk of the galaxy population. AGNs, however, are not typically accurately characterized in SAMS. It is indeed still a matter of debate how supermassive black holes evolve within the galaxy population, and physical prescriptions struggle to predict jointly within 10 per cent the galaxy and AGN population (Caplar, Lilly & Trakhtenbrot 2015, 2018; Mayer & Bonoli 2019).

Hydrodynamical simulations, coupled with sub-grid prescriptions, are able to reproduce the XLF of AGN (e.g. McAlpine et al. 2017; Biffi, Dolag & Merloni 2018). However, due to computational limitations, they span volumes that contain only small numbers of AGN and thus cannot provide the accuracy needed over the volume considered here (full sky), in particular, at the bright end of the luminosity function (e.g. Hirschmann et al. 2014, fig. 9; or Koulouridis et al. 2018, fig. 4).

The model created here is empirical, and follows the principles outlined in Croton (2009). It is constructed to reproduce accurately the observed number densities of AGN as a function of observed X-ray luminosity and redshift. Methodologically, it resembles the

¹http://www.mpe.mpg.de/comparat/eROSITA_AGN_mock/

²<https://www.cosmosim.org/cms/simulations/mdpl2>

Table 1. Light cone shells specifications. Area: sky coverage subtended (before replication). RA and Dec.: angular size (or opening angle). d (min,max): minimum and maximum comoving distance to the observer. z (min,max): minimum and maximum redshift to the observer. N snapshots: number of snapshots used. N distinct: number of distinct haloes contained in the projected simulation within the RA, Dec., and redshift boundaries that are more massive than $M_{\text{vir}} \geq 2 \times 10^{11} M_{\odot}$ (~ 100 bound particles per halo). N replications: number of replications needed to pave the full sky.

Name	Area (deg ²)	RA (deg)	Dec. (deg)	d (min,max) (Gpc)	z (min,max)	N snapshots	N distinct haloes	N replications
C1	3044	60.0	53.0	(0.000, 1.475)	(0.000, 0.365)	16	6836290	21
C2	894	30.6	29.4	(1.475, 2.795)	(0.365, 0.773)	13	11690505	91
C3	405	20.3	20.0	(2.795, 4.178)	(0.773, 1.344)	15	13938405	171
C4	227	15.1	15.0	(4.178, 5.590)	(1.344, 2.192)	16	13117165	325
C5	144	12.1	12.0	(5.590, 7.017)	(2.192, 3.559)	18	8520481	465
C6	99	10.0	9.8	(7.017, 8.454)	(3.559, 6.000)	21	2237965	703

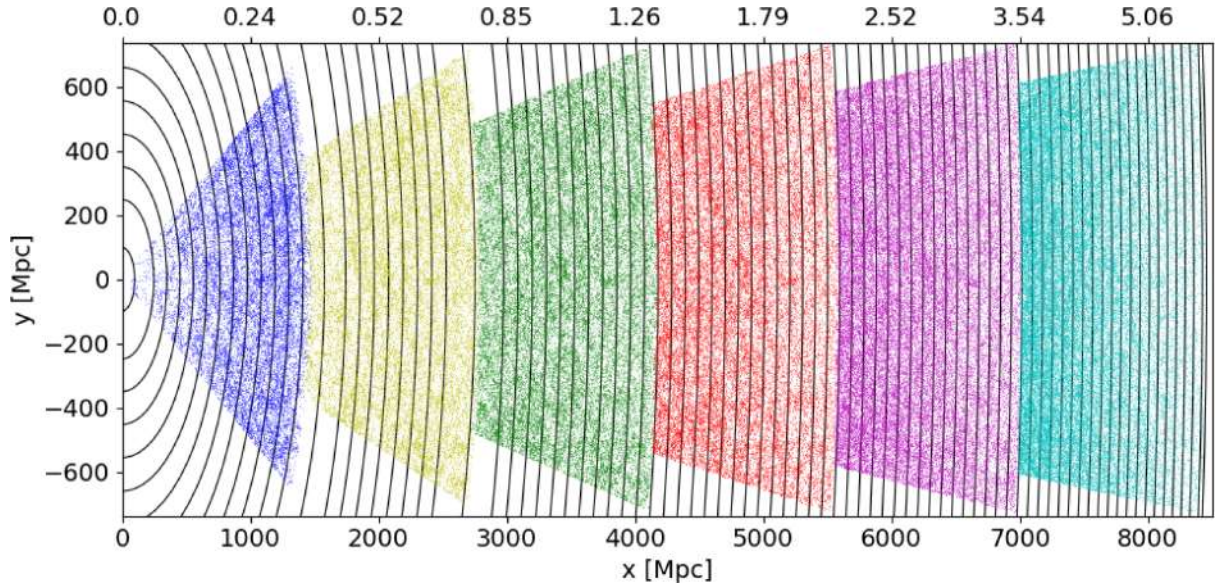


Figure 1. Layout of the box replication scheme. The six line-of-sight replications are shown in different colours, they pave up to 8500 comoving Mpc away from us, which corresponds to the redshift interval 0 to 6 (top ticks). Each dot represents a dark matter halo (only a small fraction of the haloes are represented, the thickness along the third axis is about 20 Mpc). The black lines correspond to boundaries between the shells i.e. the discrete time-steps of evolution of haloes.

halo abundance matching technique used to create high-fidelity galaxy mock catalogues, which has been extensively adopted in recent years (Kravtsov et al. 2004; Conroy, Wechsler & Kravtsov 2006; Behroozi, Conroy & Wechsler 2010; Guo et al. 2010; Trujillo-Gomez et al. 2011; Klypin et al. 2015; Favole et al. 2016; Rodríguez-Torres et al. 2016, 2017; Guo et al. 2019).

Our model goes beyond (or differs from) G18 in the following aspects:

(i) It is a generalization to all snapshots with redshift $z < 6$ re-processed into a full-sky light cone (instead of three snapshots sampling discrete redshift values 0.2, 0.75, and 1).

(ii) We moved away from sampling specific accretion rates and stellar masses from observed distributions, as in Bongiorno et al. (2016) and G18. In the full-sky simulation, the many haloes made this computationally too slow. Instead, we created a (much faster) abundance matching scheme that reproduces the luminosity and mass functions as well as the specific accretion rate distributions of super massive black holes.

(iii) Our model follows the hard XLFs from Aird et al. (2015) to better than ± 5 per cent.

(iv) An eROSITA sensitivity model for its all sky survey components is included by following the descriptions from Clerc et al. (2018a).

(v) Multiwavelength properties and classification of AGNs are computed: broad-band magnitudes (ultra-violet, optical, and infrared), plus medium resolution optical spectra.

3.1 Abundance matching model

3.1.1 Galaxy stellar masses

We use the stellar-to-halo mass relation from Moster, Naab & White (2013) to assign all distinct haloes a stellar mass M^* for their central galaxy. The logarithms of the stellar masses are drawn from a Gaussian distribution (with scatter 0.15 dex) around a mean

obtained with equation (1):

$$\bar{M}^*(M_h, z) = \frac{A(M_h, z)}{B(M_h, z) + C(M_h, z)}, \quad (1)$$

where $M_h = M_{\text{vir}}$ and

$$A(M_h, z) = 4M_h \left(0.0351 - 0.0247 \frac{z}{1+z} \right),$$

$$B(M_h, z) = \left(\frac{M_h}{10^{11.59+1.195 \frac{z}{1+z}}} \right)^{-1.376+0.826 \frac{z}{1+z}},$$

$$C(M_h, z) = \left(\frac{M_h}{10^{11.59+1.195 \frac{z}{1+z}}} \right)^{0.608+0.329 \frac{z}{1+z}}.$$

We then select haloes with $M_{\text{vir}} > 2 \times 10^{11} M_\odot$ that host galaxies with a mass $M^* > 2 \times 10^9 M_\odot$. The stellar mass function obtained is in good qualitative agreement with the literature (Ilbert et al. 2013; Davidzon et al. 2017). The obtained stellar mass function starts to turn over (i.e. become incomplete) at around $10^{10} M_\odot$ at low redshift ($z < 1$) and $10^{9.5} M_\odot$ at high redshift. In Fig. 2, we show how the stellar mass function of all mock galaxies compares with the observed stellar mass function at a similar redshift measured in the COSMOS field. Overall the mock stellar mass function is within a factor of 3 of Ilbert et al. (2013). At the high-mass end, it can differ by a factor of 20. We refrain from fine-tuning the stellar-to-halo mass relation to obtain the closest possible agreement with data. The SMF of galaxies hosting AGNs present in the mock is within a factor of 2–5 of observations.

3.1.2 AGN duty cycle

We use the AGN duty cycle values (i.e. the fraction of active galaxies at any given stellar mass and redshift) measured in Georgakakis et al. (2017), taking these from the top-right corner of each panel of their fig. 14. We linearly interpolate between the duty cycle values (0.1, 0.2, 0.3, 0.3) at redshifts (0.0, 0.75, 2.0, 6.1), respectively. This means that, e.g., at redshift 0 (2), 10 per cent (30 per cent) of the haloes are randomly chosen and designated as AGN hosts. Note that Georgakakis et al. (2017) only samples the duty cycle out to $z = 4$, so we extrapolate the $z = 4$ duty cycle value up to $z = 6$. Our model of the duty cycle is shown in Fig. 3 as a function of stellar mass, and it agrees well with the measurements of Georgakakis et al. (2017). The dependence of the duty cycle on the stellar mass for a given X-ray luminosity threshold is a consequence of the scatter in the abundance matching relation between stellar mass and X-ray luminosity (see Section 3.1.3). Fig. 2 shows how the duty cycle projects on the stellar mass function. The curves corresponding to luminosity thresholds of 10^{43} and 10^{44} based on the mock are within a factor of 2–5 of the model of Bongiorno et al. (2016). Note that we do not assume a functional form for the duty cycle as a function of stellar mass, neither do we seed black holes in the haloes.

3.1.3 Hard X-ray luminosity distributions

We adopt the AGN hard XLF from Aird et al. (2015) and specifically the LADE model therein. This provides a good (and simple) description of the XLF up to redshift 5. For our needs, we extrapolate the 2–10 keV XLF models up to redshift 6. From the hard XLF model, we draw the set of luminosities, L_X , to be attached to the AGNs in redshift shells over the full sky. We create an intermediate variable

$$\bar{M}^* = \log_{10}(L_X / [\text{erg s}^{-1}]) + N(0, \sigma) \quad (2)$$

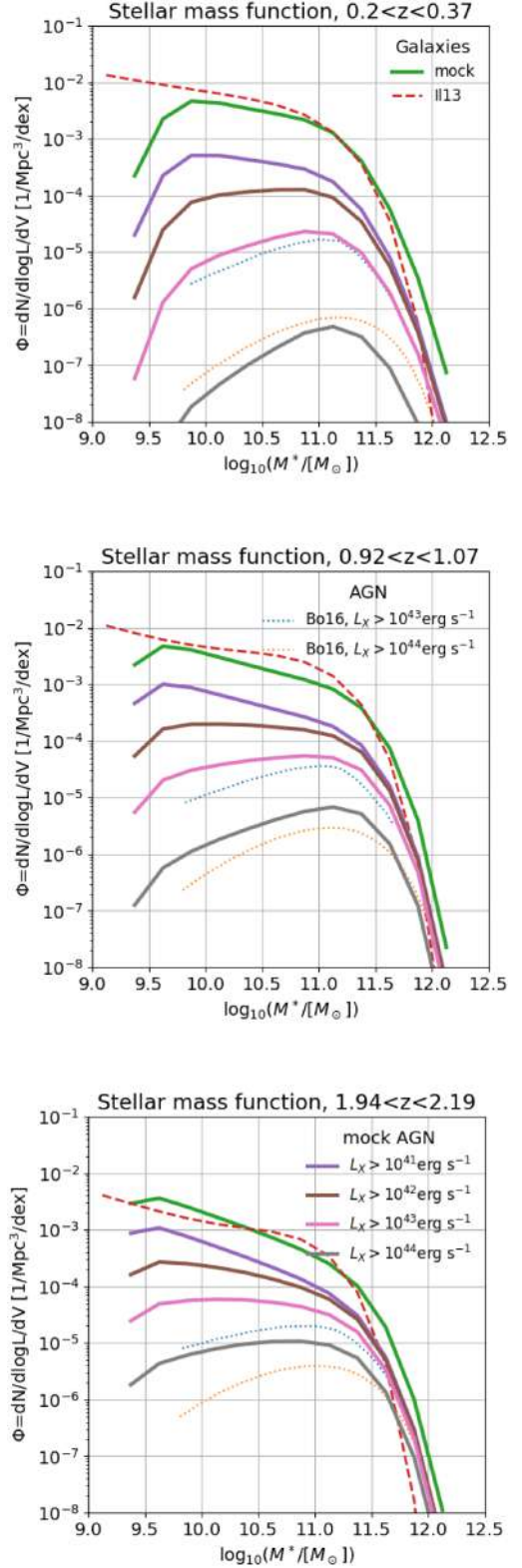


Figure 2. Stellar mass function (SMF) at redshift 0.3, 1, and 2. The SMF sampled by the mock galaxies is shown with a green solid line. It is in qualitative agreement with the SMF measured in the COSMOS survey (red dashed line, Ilbert et al. 2013). The SMFs of mock galaxies hosting AGNs (solid lines) are shown for four threshold of hard X-ray luminosity: greater than 10^{41} , 10^{42} , 10^{43} , 10^{44} erg s^{-1} . The observed SMFs of galaxies hosting AGNs (dotted lines) from Bongiorno et al. (2016) are within a factor of 2–5 from the mock.

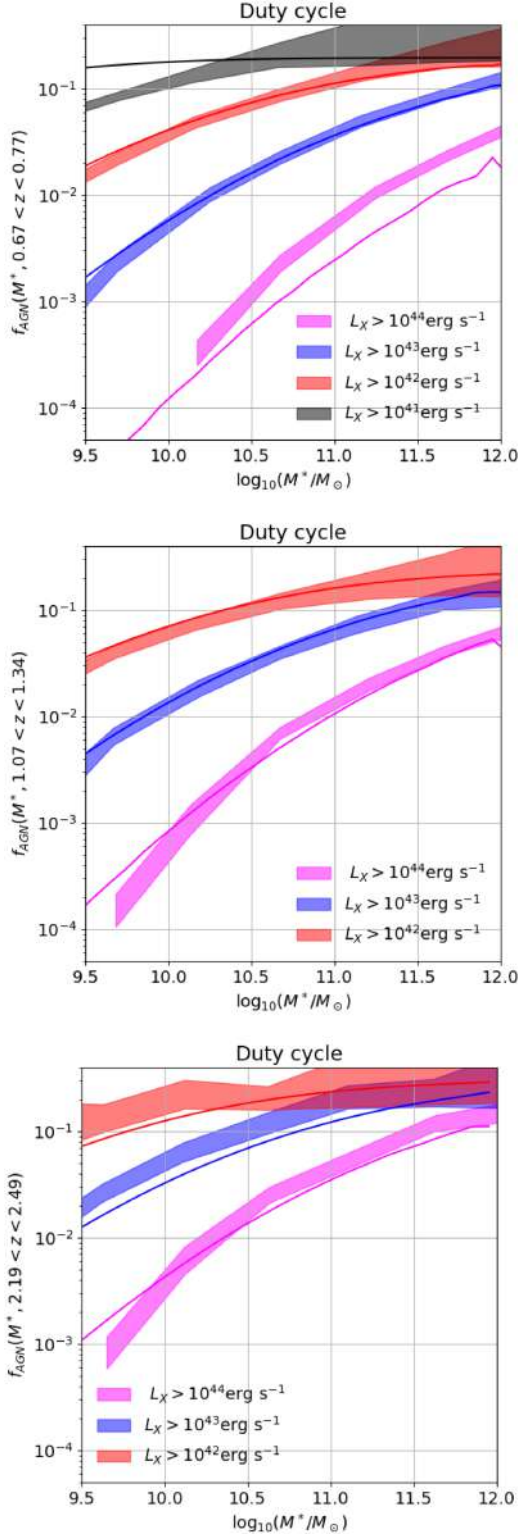


Figure 3. Fraction of active galaxies (duty cycle, DC) brighter than a given hard X-ray luminosity (2–10 keV band) threshold as a function of stellar mass at redshift 0.7, 1.2, 2.3. The DC sampled by the mock (solid lines) is in good qualitative agreement with the measurements of Georgakakis et al. (2017) at redshifts 0.75, 1.25, 2.25 (shaded areas). The maximum values at the high-mass end are in agreement by construction because they are input in the model. How the DC decreases with stellar mass for a luminosity threshold is related to the scatter in the abundance matching relation.

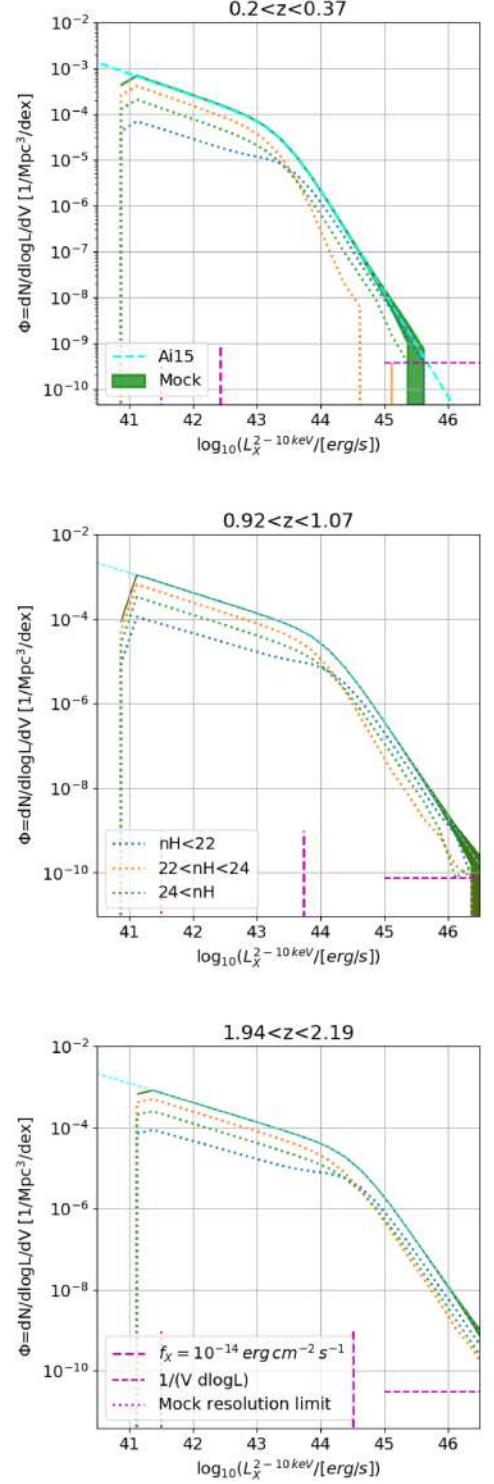


Figure 4. Hard XLF of the mock: number density versus hard X-ray luminosity (2–10 keV band) for three redshift bins: 0.25, 1.0, and 2 (green contour). This can be compared to the model from Aird et al. (2015) (blue dashed line). The decomposition of the XLF into three obscuration bins (n_H) is shown as the dotted lines. The inverse of the volume of the redshift shell is shown by the line dubbed ‘ $1/(V \, d\log L)$ ’ (horizontal magenta thin dashes). Below this limit Poisson noise dominates. The mock resolution limit is around $L_X = 10^{41.5} \text{ erg s}^{-1}$ (vertical magenta dots). The vertical magenta thick dashes give the luminosity that if emitted at the mean redshift of the bin would result in an observed hard X-ray flux of $\sim 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$.

that contains a Gaussian scatter that varies from 1.4 dex at redshift 0 to 1 dex at redshift 6 as follows:

$$\sigma(z) = 1.4 - 2z/30. \quad (3)$$

We rank order match the stellar masses to the intermediate variable: the largest M^* is connected to the largest \dot{M}^* and assigned the corresponding L_X . Adding scatter in this fashion is computationally very efficient compared with sampling from the observed bivariate distributions of stellar mass and specific accretion rate (Bongiorno et al. 2016; Georgakakis et al. 2018). This procedure guarantees the reproduction of the hard X-ray (2–10keV) luminosity function (see Fig. 4). Indeed, the mock catalogue’s XLF is within ± 5 per cent of the model. The AGN host galaxy stellar mass function for hard X-ray luminosity thresholds is shown in Fig. 2: it is in fair agreement with the model of Bongiorno et al. (2016).

3.1.4 Specific accretion rate

The ratio between hard X-ray luminosity and stellar mass gives the specific accretion rate, hereafter λ_{SAR} ($\text{erg s}^{-1} M_{\odot}^{-1} = L_X/M^*$). We compare the distributions we obtain to current estimates from Georgakakis et al. (2017) and Aird et al. (2018) in Fig. 5. The λ_{SAR} distribution is directly related to the scatter in equation (3). A large scatter, e.g. 1.5, produces a λ_{SAR} distribution that will be broad (flat), as observed in the data. A small scatter, e.g. 0.2, produces a very narrow λ_{SAR} rate distribution (i.e. a ‘light-bulb’ model for AGN activity, unlike in the data). A scatter parameter between 1 and 1.5 produces reasonable distributions λ_{SAR} and host stellar mass. It is key to reproduce the slope i.e. to sample a broad range of λ_{SAR} . The mock compares well to Georgakakis et al. (2017) and Aird et al. (2018) at low redshift at all specific accretion rates, although the observed distributions are slightly broader than in the mock at the high λ_{SAR} end. At high redshift, the slope in the mock is still in good agreement while the normalization is off. Note that, because of the large uncertainties on the measurements and the differences in coverage of λ_{SAR} between the mock and the data, the normalization of the probability distribution function (done on the mean value, only where the data are not consistent with 0) can be off by a factor of a few in the y-axis. To obtain a similar normalization, we would need to model in detail the selection functions of Georgakakis et al. (2017) and Aird et al. (2018), which is beyond the scope of this paper.

Increasing the resolution of the simulation, i.e. including lower mass galaxies further broadens the λ_{SAR} at the high λ_{SAR} end. For a galaxy that is assigned the same luminosity, its stellar mass may be smaller and its specific accretion rate higher.

3.1.5 Limitation

The main limitation of the abundance matching models come from the limited resolution of the simulation: the faint luminosity limit depends on redshift. At low redshift $z < 0.1$, we are missing some low-luminosity AGN.

3.2 Obscuration model

We create an obscuration model for the AGN similar to the ones proposed by Ricci et al. (2017) and Buchner & Bauer (2017). We add a fine redshift variation to it, following observations from Ueda et al. (2014), Aird et al. (2015), and Buchner et al. (2015). We consider three bins in the logarithm of the absorbing (neutral) column density

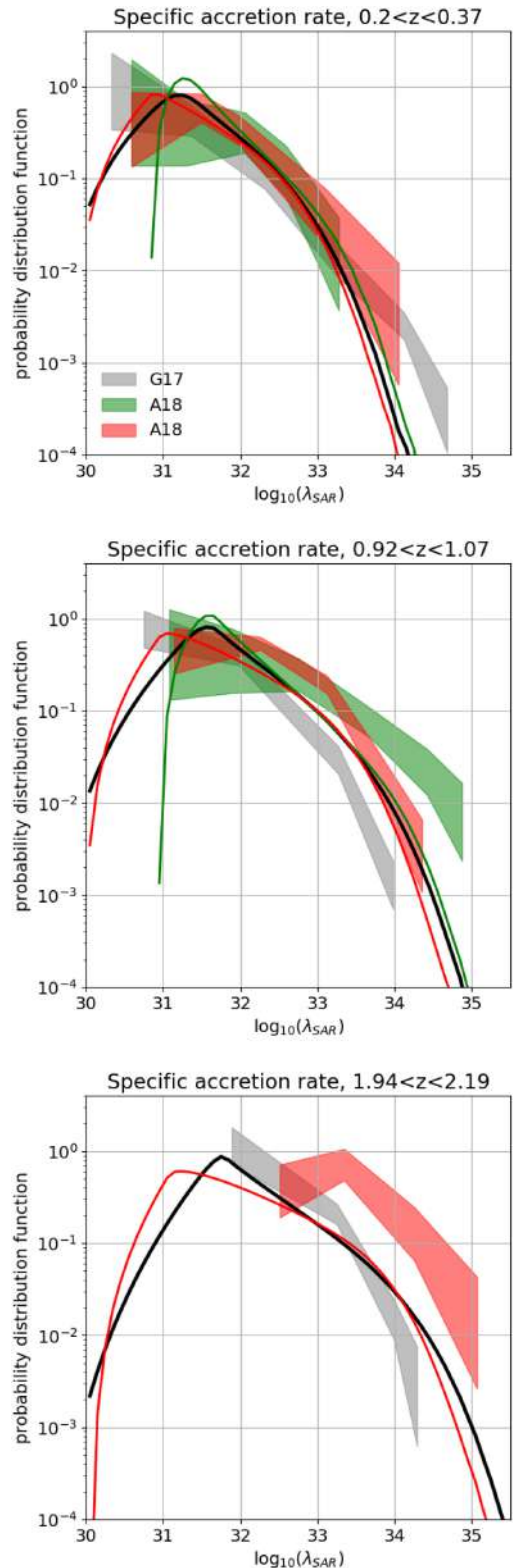


Figure 5. Specific accretion rate (λ_{SAR}) probability distribution function of the mock (solid lines) and from observations (shaded areas). The complete set [$8 < \log_{10}(M_*/M_{\odot})$] is shown in black/grey. Subsets in bins of stellar mass $9 < \log_{10}(M_*/M_{\odot}) < 10$ [$10 < \log_{10}(M_*/M_{\odot}) < 11$] are in green (red). It is key to reproduce the slope i.e. to sample a broad range of λ_{SAR} . The slope obtained in the mock compares well to the ones measured in Georgakakis et al. (2017, G17) and Aird, Coil & Georgakakis (2018, A18), although the observed distributions are slightly broader than those in the mock.

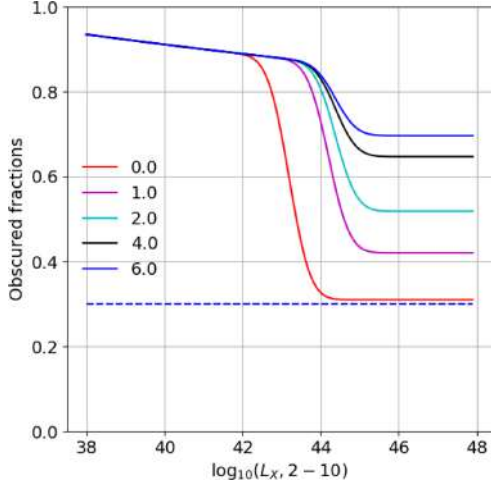


Figure 6. Model of the obscured fraction of AGN as a function of hard X-ray luminosity (2–10 keV band) at different redshifts. The blue dashed line shows the fraction of CTK objects: 0.3. The solid lines show the summed fractions of CTK plus CTN AGN at redshifts 0, 1, 2, 4, and 6.

(n_H): 20–22 unobscured, 22–24 Compton-thin obscured (CTN), 24–26 Compton-thick obscured (CTK). This model gives, as a function of hard X-ray luminosity and redshift, the fraction of AGN in each obscuration category.

The fraction of CTK AGN is fixed at 30 per cent:

$$f_{\text{thick}} = 0.3. \quad (4)$$

The fraction of CTN plus CTK AGN depends on redshift and hard X-ray luminosity :

$$f_{\text{thin+thick}}(l_x, z) = f_1(z) + (f_2(l_x) - f_1(z)) \times \dots \quad (5)$$

$$\dots \left[\frac{1}{2} + \frac{1}{2} \epsilon \left(\frac{LL(z) - l_x}{0.6} \right) \right] \quad (6)$$

where

$$l_x = \log_{10}(L_x / [\text{erg s}^{-1}]), \quad (7)$$

$$f_1(z) = f_{\text{thick}} + 0.01 + 0.4\epsilon(z/4), \quad (8)$$

$$f_2(l_x) = 0.9\sqrt{41/l_x}, \quad (9)$$

$$LL(x) = 43.2 + 1.2\epsilon(x) \quad (10)$$

and ϵ is the Gauss error function

$$\epsilon(x) = \frac{2}{\sqrt{\pi}} \int_{t=0}^{t=x} e^{-t^2} dt. \quad (11)$$

Fig. 6 shows how equation (5) evolves as a function of L_x and redshift. It is similar to that of Ricci et al. (2017); Buchner & Bauer (2017).

We run one random number per AGN, denoted \mathcal{R} . It is uniformly distributed between 0 and 1. We use it to assign the $n_H/(\text{cm}^{-2})$ values to each AGN. For each AGN, we evaluate $f_{\text{thin+thick}}(l_x, z)$ using equation (5) and compare it to the random number. If $\mathcal{R} \geq f_{\text{thin+thick}}(l_x, z)$, the AGN is classified as unobscured and we assign $\log_{10}(n_H)$ randomly in the interval 20 (included) and 22 (excluded). If $\mathcal{R} \leq f_{\text{thick}} = 0.3$, the AGN is classified as thickly obscured and

we assign $\log_{10}(n_H)$ randomly in the interval 24 (included) and 26 (excluded). Else, the AGN is thinly obscured, we assign $\log_{10}(n_H)$ randomly in the interval 22 (included) and 24 (excluded).

$$\log_{10}(n_H) \sim \text{Uniform}(20, 22) \text{ if } \mathcal{R} \geq f_{\text{thin+thick}}, \quad (12)$$

$$\log_{10}(n_H) \sim \text{Uniform}(24, 26) \text{ if } \mathcal{R} \leq f_{\text{thick}}, \quad (13)$$

$$\log_{10}(n_H) \sim \text{Uniform}(22, 24) \text{ otherwise.} \quad (14)$$

The fraction f_{abs} defined as

$$f_{\text{abs}} = \frac{N(22 < \log_{10}(n_H) < 24)}{N(20 < \log_{10}(n_H) < 24)} \quad (15)$$

thus depends on luminosity. The decomposition of the soft XLF into the different populations of unobscured, obscured, and thick obscured AGN is shown in Fig. 4. We show the obtained f_{abs} as a function of hard X-ray luminosity in Fig. 7, it is compared to observational results used to design the model (Ueda et al. 2014; Aird et al. 2015; Buchner et al. 2015). Fig. 8 shows the evolution of f_{abs} and f_{thick} with redshift for a set of luminosity thresholds. The measurement of the obscured and CTK fraction is model dependent, so, to be consistent with the XLF from Aird et al. (2015) used in the abundance matching procedure, our obscured fraction model follows their constraints up to redshift ~ 2 (see Fig 7). At higher redshifts [$z > 2$ or $\log_{10}(1+z) > 0.5$], the data do not constrain well the obscured fractions. Ueda et al. (2014), Aird et al. (2015), and Buchner et al. (2015) find CTK fractions and obscured fractions that are in tension with each other. For the obscured fraction, for simplicity, we choose not to add a component to the model to reproduce a tentative decrease in the obscured fraction at high redshift. So it may well be that our model produces too many obscured AGNs at high redshift. eROSITA data will constitute a great benchmark for the bright and high redshift end of the obscuration model. Based on observations shown in Fig. 8, the fraction of CTK AGNs could vary between 10 and 50 per cent. For simplicity, we use a constant fraction set at 30 per cent.

3.3 Soft and hard X-ray flux

To convert the hard X-ray luminosities to hard and soft X-ray fluxes, we use the obscuration model combined with a set of X-ray spectra. With the n_H values, we convert the rest-frame $L_{X, 2-10\text{keV}}$ to an observed-frame flux in the soft and hard X-ray bands: 0.5–2 and 2–10 keV. We tested two models of AGN spectra, which provide almost indistinguishable results. The first one is a Clumpy torus model with reflection, see Buchner et al. (2018). The second is the one used in Aird et al. (2015), i.e. an absorbed power law with Compton reflection and a soft scattered component at the 2 per cent level. For consistency with the XLF (also taken from Aird et al. 2015), we implemented the second set of spectra in the mock catalogue. Fig. 9 shows the combined effects of the K -correction (redshift effect) and band conversion (from hard 2–10 keV to soft 0.5–2 keV). We used them to convert the hard X-ray luminosities (2–10 keV band) to observed fluxes in the soft and the hard bands.

3.4 Galactic extinction and number counts

We include the effects of the Galactic foreground extinction following the map from HI4PI Collaboration et al. (2016). We use an average X-ray extinction law for AGN. It is computed by modelling

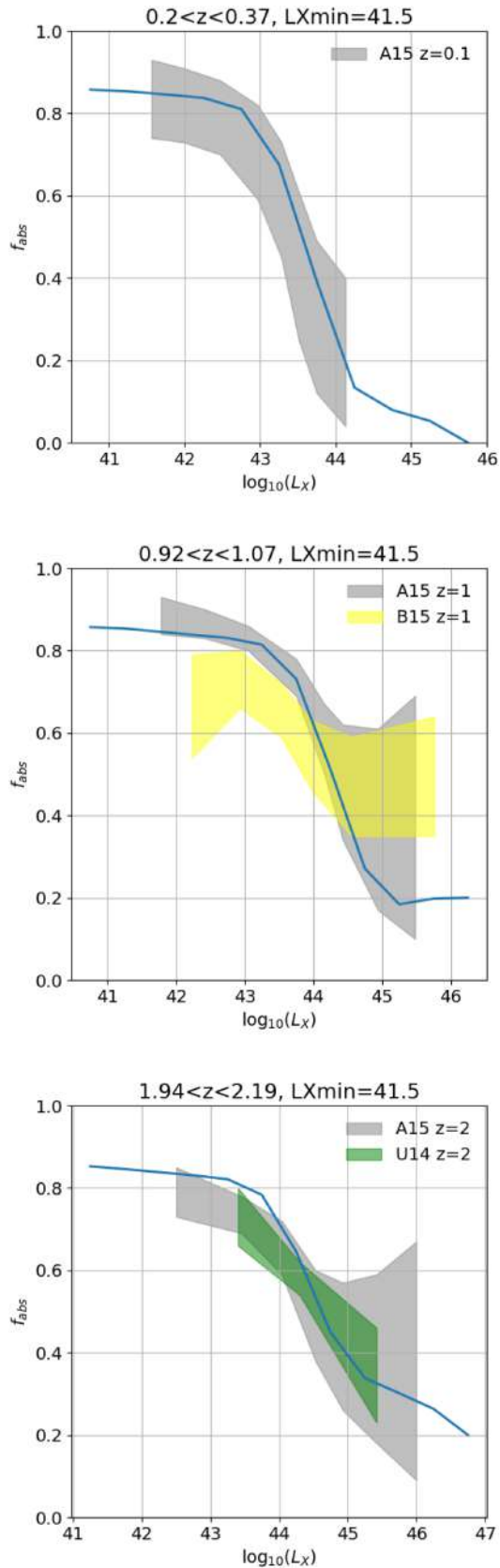


Figure 7. f_{abs} as a function of hard X-ray luminosity (2–10 keV band) at three redshifts: 0.25, 1, and 2. It is compared to its equivalent in the observations (U14, Ueda et al. 2014; A15, Aird et al. 2015; B15, Buchner et al. 2015).

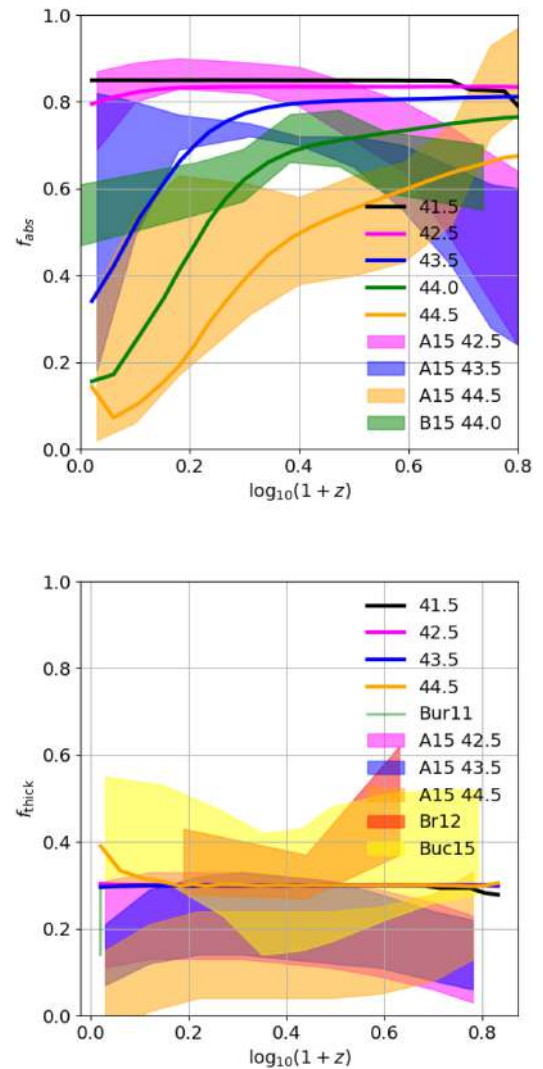


Figure 8. f_{abs} (f_{thick}) as a function of redshift for a set of hard X-ray luminosity (2–10 keV band) thresholds, top (bottom) panel.

analytically the spectra emerging through an absorber with an effective optical depth defined as the geometric mean between scattering and absorption (Yaqoob 1997); we use the absorption cross-sections of Morrison & McCammon (1983).

To determine the number of AGN as a function of flux (logN–logS relation) (Georgakakis et al. 2008; Mateos et al. 2008; Merloni et al. 2012), we consider areas with galactic latitude $|b_{\text{gal}}| > 20$ deg to avoid regions of high Galactic extinction. The logN–logS relation is shown for the hard (left-hand panel) and soft X-ray flux (right-hand panel) in Fig. 10. The mock catalogue reproduces this to within ± 20 per cent (see the right-hand panel of Fig. 10). The shape of the relation varies in the literature at the 40 per cent level, but the total number of AGNs with a flux greater than $F > 10^{-14}$ erg cm $^{-2}$ s $^{-1}$ in the mock is within a few per cent of the measurements from Georgakakis et al. (2008), Mateos et al. (2008), and Merloni et al. (2012), where the published values are in good agreement with each other. At the bright end, it is likely that the mock is missing objects due to the XLF sampling bias at high luminosities. At the faint end and at low redshift, we may be missing objects due to the resolution of the simulation.

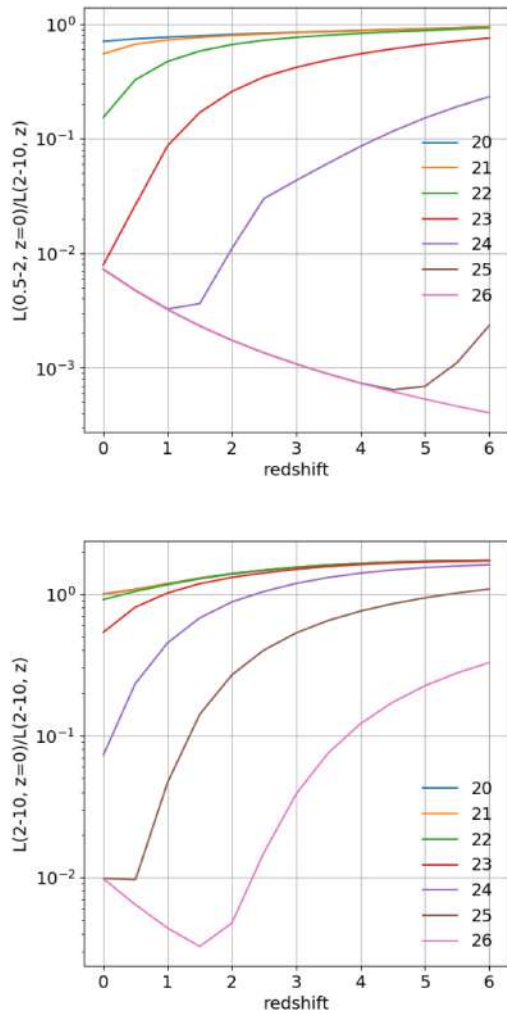


Figure 9. Conversion from the rest-frame hard X-ray luminosity (2–10 keV band) to fluxes in the soft (top panel) and hard (bottom panel) bands as a function of redshift for a discrete set of $\log_{10}(n_{\text{H}}/\text{cm}^{-2})$ values.

3.5 eROSITA sensitivity and limiting fluxes

eROSITA will scan the full sky every 6 months for a period of 4 years (Merloni et al. 2012), the data available after the first full-sky scan is named ‘eRASS:1’, after the second scan ‘eRASS:2’ and so on until the final ‘eRASS:8’, in order of increasing survey depth. We adopt here a preliminary eROSITA exposure map (Merloni et al. 2012) that assumes a simple scanning strategy. The median and mean exposure times for eRASS:8 are 1721s and 2178s. The minimum is 804s and the maximum 57 421s. We draw flux limits depending on the exposure time following Clerc et al. (2018a, appendix A). For each object, we sample from the detection probability distribution function given at its sky location to determine if it is detected or not. This detection probability function is calibrated through detailed simulation of the measurement process (see Clerc et al. 2018a). We quote flux limits for 5σ point source detection i.e. constrained by a spurious detection rate smaller than 1 in a million. For a 2 ks exposure (typical of eRASS8 on the equator), the 50 per cent completeness flux is $1.1 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$. Decreasing the significance of the detection to 3σ increases the spurious detection rate to $1/370 = 0.2$ per cent and decreases the flux limit by a factor of ~ 2 . In this analysis, we consider three different catalogues, obtained with three different detection limits for three different phases of the

survey: an ‘early’ survey catalogue compiled from 5σ detections after three full-sky scans (named eRASS3 hereafter), a ‘secure’ 5σ detection catalogue after eight full-sky scans (eRASS8), and 3σ detection catalogue after eight full-sky scans (SNR3). For the rest of the paper, unless mentioned otherwise, we adopt the ‘eRASS8’ as the baseline case (the other cases are presented in the Appendix).

3.6 Optical and X-ray AGN classification

In order to associate optical types to X-ray detected AGN in a meaningful and realistic way, we use the ‘obscuration-matrix’ derived from Merloni et al. (2014). This describes, on the basis of the *XMM-Newton* X-ray-selected AGN catalogue in the COSMOS field, the relative fraction of X-ray or optically obscured AGN by splitting them into four categories: unobscured in both optical and X-rays (‘11’), X-ray obscured and optically unobscured (‘12’), X-ray unobscured and optically obscured (‘21’), obscured in both X-rays and optical (‘22’). Here, optically ‘obscured’ means that no broad line or blue accretion disc continuum (big blue bump) is detected in the optical spectrum. We parameterize the boundaries as a function of X-ray luminosity between the 21 and the 11 as well as between the 22 and 12 classes (visible in fig. 12 of Merloni et al. 2014) with a Gauss error function (ϵ defined in equation (11):

$$f_{r_{\text{split}}}(l_x) = \left[\frac{1}{2} + \frac{1}{2} \epsilon \left(\frac{44 - l_x}{0.9} \right) \right] 0.69 + 0.26, \quad (16)$$

where l_x and ϵ are defined in equations (7) and (11). This boundary further splits the X-ray obscured–unobscured classification.

We compute the obscured fraction of AGN (f_{r_o}) as a function of X-ray luminosity using the boundary $\log_{10}(n_{\text{H}}) > 22$. $f_{r_o} - f_{r_{\text{split}}}$ gives the fraction of optically unobscured AGN among the X-ray obscured population: type 12. We select them randomly among the X-ray obscured [$\log_{10}(n_{\text{H}}) > 22$] sources to match the fraction needed. $f_{r_{\text{split}}} - f_{r_o}$ gives then the fraction of optically obscured AGN among the X-ray unobscured: type 21. We select these randomly among the X-ray unobscured [$\log_{10}(n_{\text{H}}) < 22$] to match the fraction needed. The X-ray obscured (unobscured) AGNs that were not assigned to type 12 (21) are assigned type 22 (11). Among the X-ray obscured AGNs ($n_{\text{H}} > 22$), most become type 22 and some type 12 (at high luminosity). Fig. 11 shows the classification obtained as a function of L_X at three redshifts: 0.3, 1, and 2.

3.7 Spectral energy distribution in the ultraviolet, optical, and infrared

Determining redshifts for the eROSITA AGN is key for most scientific applications. A redshift measurement with an accuracy of ~ 1 per cent and outlier fraction of 5 per cent is possible using a set of broad-band photometric magnitudes (i.e. ‘photo-z’ Salvato et al. 2011; Salvato, Ilbert & Hoyle 2018a; Salvato et al. 2018b). Photo-z of this accuracy are sufficient for luminosity function studies, for example, and the SEDs more generally can provide information on classification. A redshift measurement with accuracy 0.05 per cent and an outlier fraction smaller than 1 per cent is feasible using spectroscopy (i.e. ‘spec-z’ Myers et al. 2015; Dawson et al. 2016; Dwelly et al. 2017). Spec-z allow studies of the clustering and of spectral line properties, including more accurate and details (sub)classifications. To emulate photo-z and spec-z, we create an empirical description of AGNs in the UV, optical, and IR.

In Section 3.7.1, we compute the SDSS *r*-band magnitude for each AGN. Then in Section 3.7.2, we link each AGN to a spectral

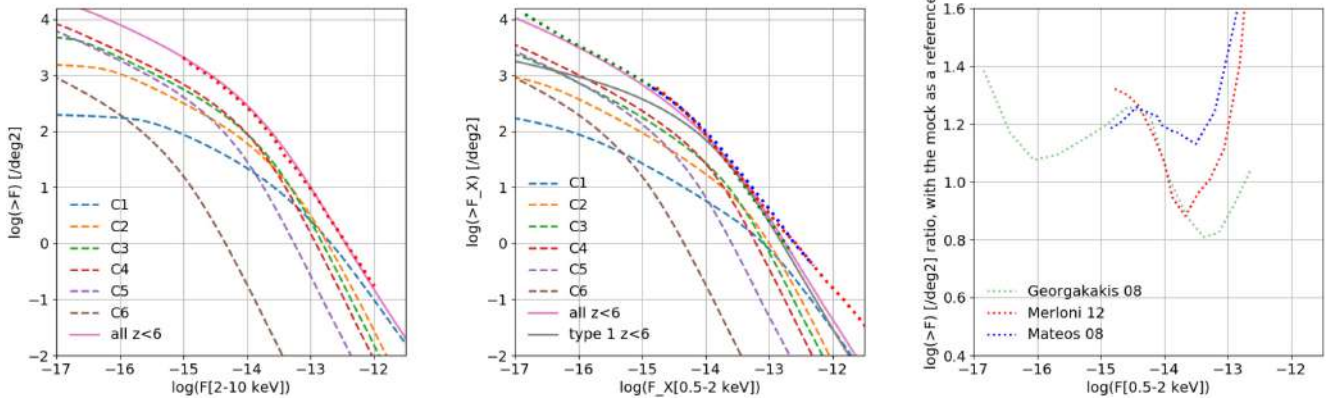


Figure 10. $\log N$ – $\log S$: AGN cumulative number density as a function of observed X-ray flux as in the mock catalogue between redshift 0 and 6 (pink solid line). The left (middle) panel shows the $\log N$ – $\log S$ relation for the hard 2–10 keV band (soft, 0.5–2 keV band) X-ray flux, divided into six redshift bins corresponding to the light-cone shells (dashed lines) with C1: $0 < z < 0.36$; C2: $0.36 < z < 0.77$; C3: $0.77 < z < 1.34$; C4: $1.34 < z < 2.19$; C5: $2.19 < z < 3.55$; C6: $3.55 < z < 6$. This is compared to Georgakakis et al. (2008), Mateos et al. (2008), Merloni et al. (2012) (green, blue, and red dotted lines). In the right-hand panel is shown the ratio of these observed curves to the one from the mock for the soft X-ray flux. Around the flux limit of eROSITA, $\log_{10}(F_X^{0.5-2}) = -14$, the agreement is good, which is the key issue for the validity of the mock. At fainter fluxes, the mock is missing about 20 per cent of the sources. At the bright end because of the way the sampling is done and the finite volume of the simulation shells, we miss the brightest objects. The grey solid curve on the middle panel shows the $\log N$ – $\log S$ relation when restricted to type-1 AGNs.

energy distribution extending from the UV to the IR to compute a set of broad-band magnitudes. Finally in Section 3.7.3, we link each AGN to a medium resolution optical spectrum to emulate spectroscopic observations.

3.7.1 Optical SDSS r -band magnitude

In this section, we describe the method we use to compute the SDSS r -band magnitudes (Fukugita et al. 1996). To do this, we used the observed F_X/F_{opt} ratio or more precisely the relation between soft X-ray flux and r -band magnitude from existing surveys to calibrate the relationship and its scatter. The r -band magnitude can then also be used to normalize the overall SED. The F_X/F_{opt} relationship could in principle depend on the optical type, because of the differential effects of obscuration on the X-ray and optical fluxes, and the differing contributions of the host galaxy.

We therefore examine this relation for three optical classes of AGNs:

- (i) Optical spectroscopic type 1 corresponding to type 11 or type 12, with quasar-like spectra.
- (ii) Optical spectroscopic type 2 corresponding to type 22 or type 21, with type2-like spectra.
- (iii) The ‘elusive’ AGN, or XBONGS (X-ray bright, optical normal galaxies) (Comastri et al. 2002; Menzel et al. 2016); we call these type-3 AGN. We take a random 20 per cent of the optical spectroscopic type-2 AGNs to be represented by this class. Their spectrum is galaxy like.

We performed a fit to the ratio of the 0.5–2 keV flux to the optical r -band magnitude, using the CDFS, Stripe82X, COSMOS, and 2RXs catalogues (Hsu et al. 2014; Fliri & Trujillo 2016; Ananna et al. 2017; Marchesi et al. 2016; Salvato et al. 2018b). In these catalogues, the AGNs are sorted according to optical type 1 (unobscured), 2 (obscured), or 3 (galaxy like), as defined by the best-fitting SED template or spectroscopic information. Note that the classification is not perfect and the three classes are to some extent contaminating each other. Also, the total number of type 2 and 3 is small with respect to the total number of type 1. So we fit

one relation for type 1 and one for type 2 and 3 together. We find that a mean relation works for all populations:

$$\bar{r}(\log_{10} F_X) = -2 \log_{10}(F_X) - 7. \quad (17)$$

The scatter around the relation is well approximated by a Gaussian with unity scatter for all populations:

$$\sigma = \mathcal{N}(0.0, 1.0). \quad (18)$$

Fig. 12 shows the data considered in the r magnitude versus soft X-ray flux plane. It shows that a single relation accounts well for all types. It shows a hint of trend with redshift, more data are required to confirm this trend.

Current surveys (e.g. COSMOS, Stripe82, and CDFS) lack sufficient number of AGN to constrain an eventual joint fit of the X-ray and optical luminosity functions.

Note that current data do not contain sufficient number of AGN to constrain an eventual joint fit of the X-ray and optical luminosity functions. A complete discussion on the possible universality of this relation is left for future studies. Indeed, future surveys such as eROSITA, 4MOST, and the Maunakea Spectroscopic Explorer (Merloni et al. 2019; Richard et al. 2019; The MSE Science Team et al. 2019) will hopefully settle this.

3.7.2 Spectral energy distribution and broad-band photometry

To test the algorithms that identify the AGN counterparts in the UV, optical, and IR and measure the photo- z , we need to generate mock magnitudes corresponding to those that will be available for the data sets that will be used in reality. We use spectral energy distributions (SED) models from Ciesla et al. (2015) to deduce UV, optical, and near IR magnitudes for all AGN simulated in the mock. The SEDs are shown in Fig. 13. We use a model SED for each spectroscopic optical type: 1, 2, 3. The following band passes are simulated: GALEX-NUV, GALEX-FUV, HSC-g, HSC-r, HSC-i, HSC-z, HSC-y, UKIDSS-J, UKIDSS-H, UKIDSS-K, WISE-W1, WISE-W2 (Martin et al. 2005; Hewett et al. 2006; Wright et al. 2010; Kawanomoto et al. 2018). These surveys are of particular interest because they are located in the performance

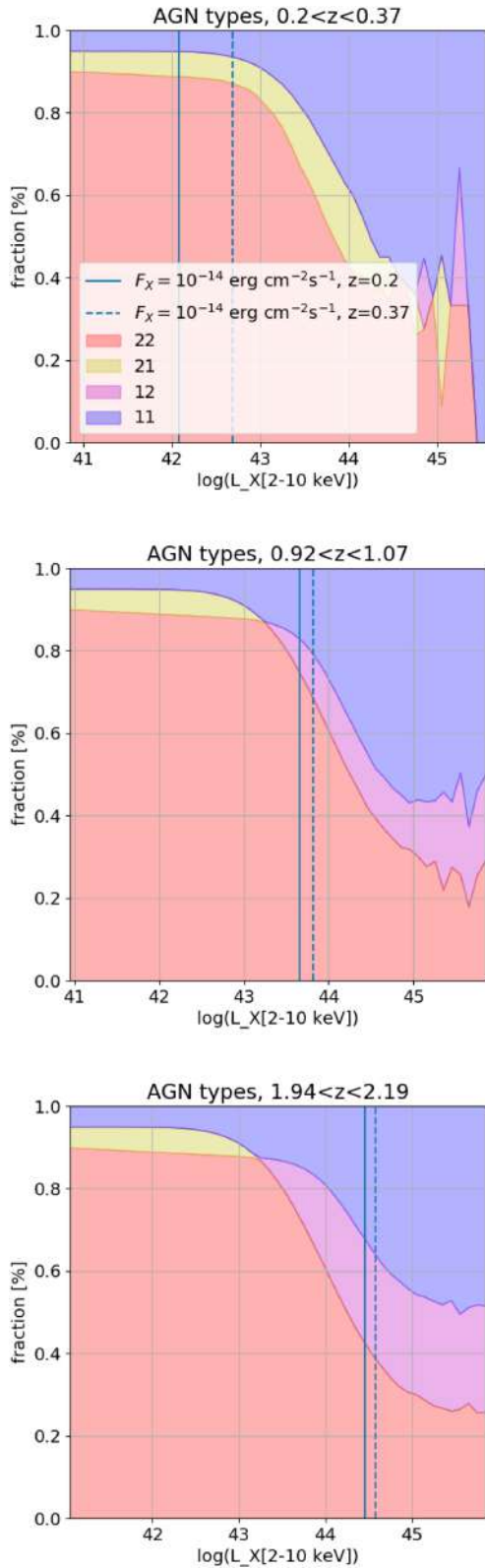


Figure 11. Fractions of AGNs as a function of hard X-ray luminosity (2–10 keV band) in each category: unobscured in both optical and X-rays (‘11’), X-ray obscured and optically unobscured (‘12’), X-ray unobscured and optically obscured (‘21’), obscured in both X-rays and optical (‘22’). The vertical lines show the luminosity corresponding to the flux of 10^{-14} erg cm $^{-2}$ s $^{-1}$ (close to the eROSITA detection limit) at the boundaries of the redshift interval considered.

verification equatorial commissioning field of eROSITA. We use the model SED, rescaled to the SDSS r -band magnitude, and create a magnitude array and its uncertainty to mimic the different depths of the GALEX, HSC, VIKING, and WISE surveys (Morrissey et al. 2007; Edge et al. 2013; Aihara et al. 2018; Dey et al. 2019). To mimic the depth, we fit with a polynomial of second or third degree the magnitude – error relation in each of these surveys in a representative region. Table 2 gives the coefficients of the polynomial

$$\log_{10}(\sigma_m(m)) = \sum_i c_i m^i \quad (19)$$

that links an AB magnitude, m , to its uncertainty σ_m . For individual AGN, we draw uncertainties around this relation with a Gaussian scatter of 0.158 dex.

3.7.3 Optical medium resolution spectra

We link each AGN in the mock to a representative medium resolution ($R \sim 3000$) optical spectrum by stacking SDSS spectra (Abolfathi et al. 2018; Pâris et al. 2018) with the archetype method Zhu (2016) to represent each of the three types as a function of redshift.

We create a set of medium resolution optical spectra for the type 1, 2, 3 AGNs as well as for red, elliptical galaxies and for blue, star-forming galaxies.

For the type-1 AGNs, there exist a large number of optical spectra, and the stacking method permits the construction of high-quality stacks that are almost noise free (see Fig. B1).

For the type-2 AGNs, the number of available spectra is small, so that the stacks have noise, in particular at redshift higher than 1 (see Fig. B2).

For the type-3 AGNs, we use the elliptical, red galaxy spectrum model (see Fig. B3).

To construct galaxy archetypes, we use SDSS spectra as selected in Abazajian et al. (2009), Dawson et al. (2013), and Prakash et al. (2016) for the elliptical galaxies (Fig. B3) and as in Comparat et al. (2013, 2016), and Raichoor et al. (2017) for the star-forming galaxies (Fig. B4).

We use the Planck dust maps with the DUSTMAP package³ to redden the optical SDSS r -band magnitude as well as the spectra. We use the EXTINCTION⁴ package to attenuate magnitudes and spectra with the Fitzpatrick (1999) extinction law (the r LF comparison discussed above is done on de-reddened magnitudes). This is used below for spectroscopic exposure time calculation.

4 RESULTS: THE PREDICTED EROSITA AGN POPULATION

The main application of the model is the prediction of the distribution of eROSITA AGN on the sky in redshift and magnitude. These distributions show how the eROSITA all-sky survey will sample our past light cone in an unprecedented fashion.

The total expected number of AGNs is 2.3 (0.9, 3.5) million for the eRASS8 scenario (eRASS3, SNR3); previous estimates were around 3 for eRASS8. The discrepancy is mainly due to previous sensitivity maps (Merloni et al. 2012; Kolodzig et al. 2013) that were more optimistic than the more recent one from Clerc et al. (2018b).

³<http://dustmaps.readthedocs.io/en/latest/index.html>

⁴<https://extinction.readthedocs.io>

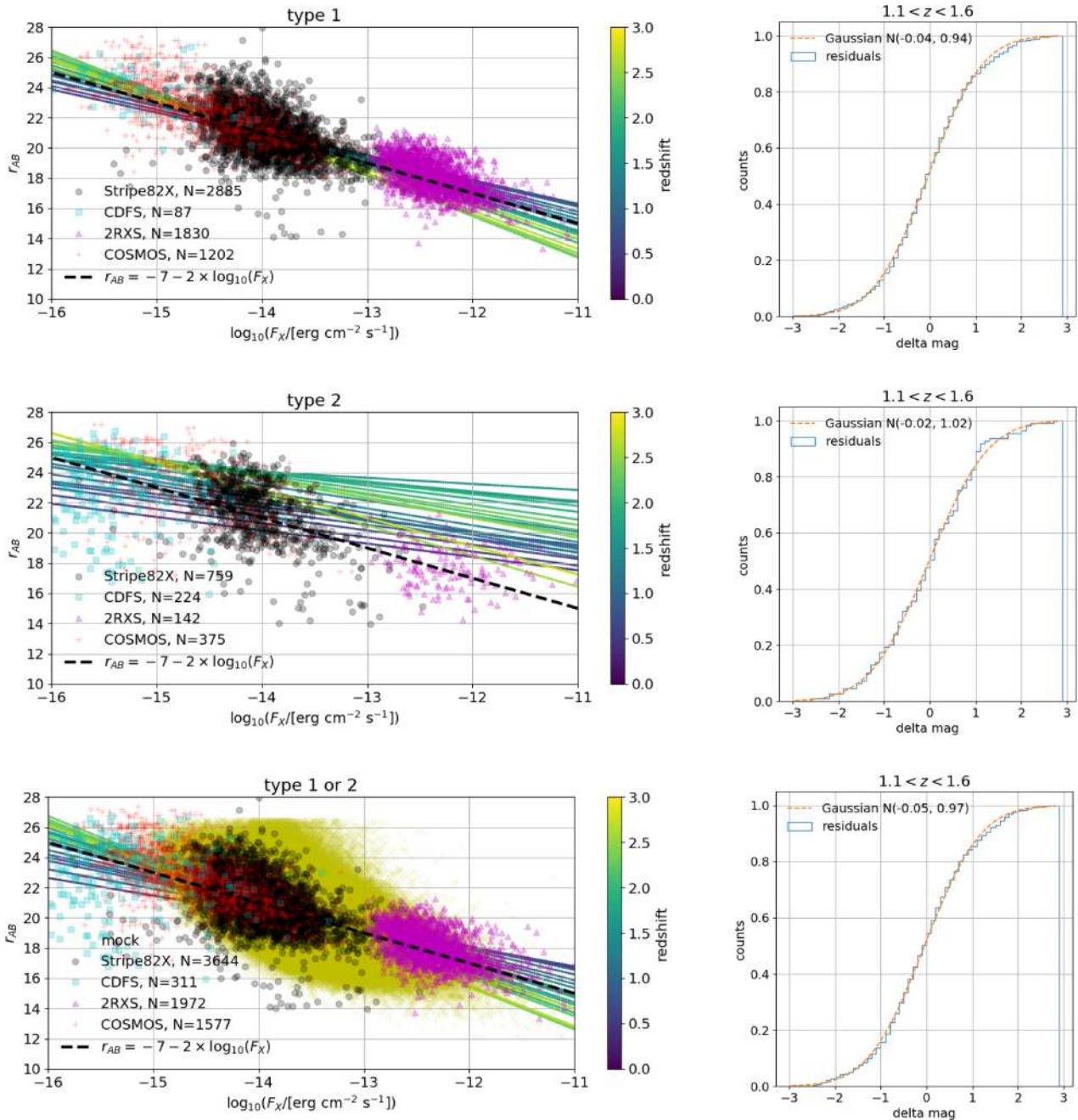


Figure 12. r magnitude versus soft X-ray flux (left) and cumulative distribution of residuals around the model (right) for type-1 AGN (top), type-2 AGN (middle), and both (bottom) in the redshift bin $1.1 < z < 1.6$. The CDFS, Stripe82X, COSMOS, and 2RXs catalogues are shown in cyan circles, black circles, red crosses, magenta triangles (Hsu et al. 2014; Fliri & Trujillo 2016; Marchesi et al. 2016; Ananna et al. 2017; Salvato et al. 2018b). The yellow crosses in the bottom left-hand panel show the eRASS8 mock. The coloured lines show the best fits as a function of redshift. The black dashed line shows the relation adopted.

As the instrumental background for eROSITA at the L2 point is highly uncertain, the exact sensitivity as a function of exposure time is not easy to predict. Having the three scenarios is thus handy to see how total number evolve with total exposure time assumed (3/8 exposure time ratio between eRASS3 and eRASS8 scenarios) and detection significance (3σ and 5σ detection in the SNR3 and eRASS8 scenarios).

The expected variation in AGN number density as a function of right ascension and declination in a ‘car’ projection is shown in Fig. 14 (Fig. B5 shows the same projection for the eRASS3 and

SNR3 scenarios). Away from the poles and the milky way, eROSITA will typically detect between 60 and 100 AGN per square degree. The ecliptic poles, where the eROSITA exposure is the longest, are the most densely populated regions with densities around 250 deg^{-2} in the eRASS8 scenario. For the eROSITA telescopes, the angular half energy width at 1.5 keV is ≈ 16 arc seconds on-axis, and degrades to an arc minute off-axis. Source confusion thus influences at source densities above $1 \text{ arc minute}^{-2} \sim 1100 \text{ deg}^{-2}$; thus, source confusion can safely be ignored for most areas to be surveyed by eROSITA. Note that the maximum density (in the extra-Galactic

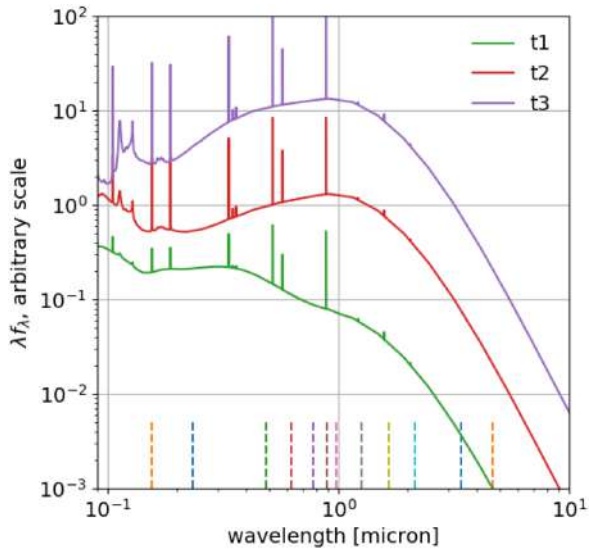


Figure 13. Flux versus wavelength (microns) for the three spectral energy distributions used to compute AGN broad-band magnitudes. The curves t1, t2, and t3 represent the optical classes described in the text: type 1, 2, and 3. The vertical dashed lines show the mean wavelengths of the broad-band filters used to compute magnitudes, from left to right: GALEX-NUV, GALEX-FUV, HSC-g, HSC-r, HSC-i, HSC-z, HSC-y, UKIDSS-J, UKIDSS-H, UKIDSS-K, WISE-W1, WISE-W2.

sky) is reached in the SNR3 scenario with a maximum of 400 deg^{-2} in the poles, so that in this specific case and region source confusion may have an impact.

We give the average redshift distribution of the AGNs for scenarios eRASS8, eRASS3, and SNR3 in the Tables 3, A1, and A2. They give the average redshift distributions over the extra-Galactic sky ($27k \text{ deg}^2$) and at the poles. We predict an average detection rate of $\sim 74 \text{ AGN deg}^{-2}$ (218) on the extra-Galactic sky (pole). We find good agreement with previous predictions concerning the shape of the redshift distribution (Merloni et al. 2012; Kolodzig et al. 2013). The redshift distribution is illustrated in Figs 15, B6, and B7 for the scenarios eRASS8, eRASS3, and SNR3. These show how the poles compare to the general extra-Galactic sky area. The redshift distributions are shown in four magnitude ranges that are appropriate for the facilities that will be used for optical follow-up (see Section 4.2).

4.1 Optical luminosity function

With the SEDs described above, we obtain a set of optical AGN LFs. They have a double power-law shape like the XLF. As shown in Fig. 16, our predicted optical LF agrees within a factor of 2 with the *i*-band LF of quasars (Croton 2009; Ross et al. 2013) and with the *g*-band LF of a complete census of optically selected AGN (Palanque-Delabrouille et al. 2016; Caditz 2017; PLE model). The observer-frame *g*- and *i*-band luminosity functions are shown in Fig. 16. We show the complete model, limited by the resolution of the simulation (grey), and the eRASS8 mock (red), limited by the eROSITA selection function. We compare the predicted optical type-1 AGN number counts as a function of the observed *r*-band magnitude in four broad redshifts bins ($0 < z < 1$, $1 < z < 2$, $2 < z < 3$, $3 < z < 4$) with the observations from Palanque-Delabrouille et al. (2016) and Caditz (2017, table 7). The number of AGNs predicted by the mock is within ± 40 per cent of the model

prediction in the range $0 < z < 3$, $19 < r < 23$ (see Fig. 17). Outside of this range, the discrepancy can reach a factor of 2. Finally, Fig. 18 shows the cumulative number density projected on the sky ($\log N - \log R$) of a set of X-ray-selected AGN samples ($F_X^{0.5-2} > 10^{-14}$, 2×10^{-14} , 2×10^{-15} , 6×10^{-15} , 8×10^{-16} , $7 \times 10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1}$) as obtained from the mock catalogue. It is compared to the Stripe82X catalogue from Ananna et al. (2017) limited to a flux 6×10^{-15} . We find a good agreement. Based on these three tests, optical luminosity function, observed magnitude number density as a function of redshift, and $\log N - \log R$ for X-ray selected AGNs, we deduce that the model is reliable to describe both the X-ray and the optical properties of the AGNs.

4.2 Planning spectroscopic follow-up of eROSITA AGN

Redshift measurement for eROSITA AGN is key for most scientific applications. Ground-based multiobject (typically fibre-fed) spectrographs are suited infrastructures to measure redshifts for large number of sources over wide areas of the sky. Based on our mock catalogue predictions (calibrated on observed multiwavelength data), the optical magnitude range of the eROSITA X-ray-selected AGN is broad, thus one needs a set of telescopes to acquire spectroscopic redshifts for all AGNs in an optimally efficient way. 1-m class infrastructure can measure redshifts up to magnitude 17–18 (e.g. TAIPAN; da Cunha et al. 2017), where cross-talk effects could negatively affect larger aperture telescopes (indeed, SDSS/BOSS spectroscopic surveys had been limited to objects fainter than 17th magnitude); 2-m class telescope are optimally suited to follow up systems up to 20–21 mag (e.g. SDSS; Bolton et al. 2012), while 4-m class telescope could measure redshift up to magnitude 22–23 (e.g. 4MOST, DESI; de Jong et al. 2012; DESI Collaboration et al. 2016), and 8-m class telescopes up to 25–26 (e.g. VIMOS, MUSE, MSE; Scodreggio et al. 2005; Bacon et al. 2015; McConnachie et al. 2016).

In order to determine an ideal strategy for the systematic spectroscopic follow-up of eROSITA AGN, we first divide the full sky in two regions: the ecliptic pole regions, containing AGN with ecliptic latitude above 80 or below -80 deg , which are deeply exposed by the SRG/eROSITA survey scanning strategy, and the extra-Galactic region (‘Xgal’) defined by a Galactic latitude cut at $|b_{\text{gal}}| > 20^\circ$ but where the poles are removed. Using the medium resolution optical templates, the SDSS *r*-band magnitude and the 4MOST exposure time calculator (Dwelly et al. 2016, version of December 2019), we estimate the exposure time needed (expressed in fibre-hours) for each source to reach a median signal-to-noise ratio of 2.5 per Angstrom, which should yield a success rate in measuring redshifts above 95 per cent. For simplicity, we then scaled this to a corresponding exposure time with a 1-, 2-, and 8-m class telescopes by dividing or multiplying exposure times by the ratio of the collecting area. Table 4 (A3, A4) shows the results of the ETC calculation for the eRASS8 (eRASS3, SNR3) scenario. About 2.3 million fibre-hours split between these infrastructures would be sufficient to complete the full spectroscopic identification of all eRASS8 AGN. Table 4 (A3, A4) also shows the average fibre-hours density needed, i.e. the product of the average number of targets per unit area times the average exposure time (in hours). In the extra-Galactic area, the average density of fibre-hours required is on the low side: $20\text{--}50 \text{ deg}^{-2}$. In the deep ecliptic-pole regions, the density of fibre-hours required is higher: $100\text{--}400 \text{ deg}^{-2}$. The fibre-hour densities required are within reach of current or near-future state-of-the-art multiobjects spectrograph instruments, such as

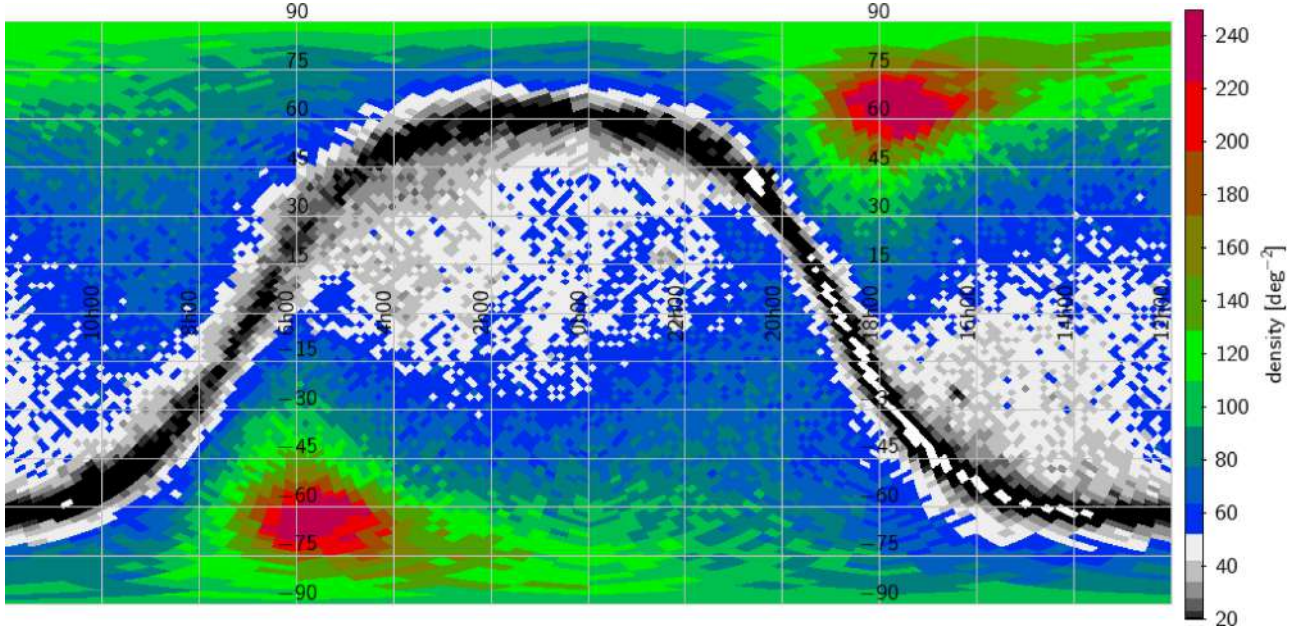


Figure 14. Distribution on the sky of the expected number density of X-ray AGNs after the fiducial eRASS8 observation plan, i.e. after exposure map and Galactic extinction is taken into account (number counts averaged on 3.35 deg^2 pixels).

Table 2. Polynomial coefficients of the magnitude error relation given in equation (19). $\bar{\lambda}$ is the mean wavelength of the broad-band filter in microns.

Band	$\bar{\lambda}$	Polynomial coefficients		
		c_2	c_1	c_0
GALEX-NUV	0.234	$7.24\text{e-}03$	$-4.28\text{e-}04$	-5.13
GALEX-FUV	0.155	$6.39\text{e-}03$	$-2.43\text{e-}04$	-4.67
HSC-g	0.485		0.337	-9.727
HSC-r	0.627		0.346	-9.788
HSC-i	0.777		0.353	-9.905
HSC-z	0.892		0.353	-9.646
HSC-y	0.979		0.341	-9.075
UKIDSS-J	1.256	0.009	-0.010	-4.611
UKIDSS-H	1.651	0.012	-0.091	-3.910
UKIDSS-K	2.153	0.014	-0.124	-3.521
WISE-W1	3.4		0.394	-9.129
WISE-W2	4.652		0.396	-8.797

- (i) 1-m class: TAIPAN ($25 \text{ fibres deg}^{-2}$, 150 fibres),
- (ii) 2-m class: SDSS ($145 \text{ fibre deg}^{-2}$, 1000 fibres),
- (iii) 4-m class: 4MOST ($400 \text{ fibres deg}^{-2}$, 1600 fibres),
- (iv) 8-m class: MOONS ($7200 \text{ fibre deg}^{-2}$, 1000 fibres).

The fibre-hour densities required are well below availability, so the eROSITA AGN follow-up can be carried out next to other programmes. At the time of writing, large spectroscopic follow-up campaigns of the eROSITA AGN population are planned for SDSS-V (Kollmeier et al. 2017) and 4MOST (de Jong et al. 2012; Merloni et al. 2019).

4.3 Expected uncertainty in the XLF

Assuming that we will have measured spectroscopically 90 per cent of the redshifts of all AGNs with magnitude $r < 23.5$, we predict the expected fractional uncertainty in the luminosity function (see Fig. 19). At low redshift, per cent precision will be reached over a wide range of luminosities and will lead to a strong anchor of the XLF models. At intermediate redshift $1 < z < 3$, the uncertainty will be around the per cent level at luminosities of $\log L_X$ 44–45 (i.e. just above the knee of the XLF), an improvement of more than one order of magnitude with respect to previous studies (which constrained it at the 20–50 per cent level) (Aird et al. 2015; Miyaji et al. 2015). This will strongly constrain the evolution of the bright end of the XLF.

4.4 Clustering on large scales, and the baryon acoustic oscillation standard ruler

The upcoming combination of eROSITA+SDSS-V+4MOST will increase the size of the current largest collection of X-ray-selected AGNs with spectroscopic redshift measurements by a factor of 100. This will allow precise clustering measurements as a function of redshift. eROSITA is unique in terms of target selection, as it spans a very wide redshift range.

Based on the redshift distribution, we predict the quality of clustering measurements on large scales and the precision with which one could measure the baryon acoustic oscillation standard ruler (Eisenstein & Hu 1998; Seo & Eisenstein 2003, 2007). Such a calculation requires two more inputs as a function of redshift: the bias and the redshift uncertainty. We measure the clustering amplitude of X-ray AGN in the mock catalogue as a function of redshift and find that the description of the bias evolution with redshift is in agreement with equation (12) of Chehade et al. (2016). We assume redshift uncertainties to be constant $\sigma_z = 0.005(1 + z)$, which is slightly more conservative

Table 3. Number density of AGN as a function of redshift for the eRASS8 scenario for different optical r -band magnitude cuts ($12 < r < 18$, $18 < r < 21$, $21 < r < 22.8$, $22.8 < r < 25$, type 1 $r < 25$, all $r < 25$). The first two columns give the boundaries of each redshift bins. All other columns give $\frac{dN_{\text{AGN}}}{\text{deg}^2}$ the average number of AGN per square degree in that redshift bin for the extra-Galactic sky (x) and for the poles (p). This was computed on an area of $27\,143$ (626.7) deg^2 for the extra-Galactic sky (poles), namely $|b_{\text{gal}}| > 20^\circ$ ($|b_{\text{ecl}}| > 80^\circ$).

z bin		$12 < r < 18$		$18 < r < 21$		$21 < r < 22.8$		$22.8 < r < 25$		Type 1		All	
Min	Max	x	p	x	p	x	p	x	p	x	p	x	p
0.0	0.4	0.3148	0.2585	3.9610	5.4617	1.8695	6.0537	0.2206	1.7041	2.7738	4.6623	6.3681	13.5338
0.4	0.8	0.3368	0.3271	9.0241	13.4732	5.1919	17.7606	0.6465	5.3947	11.5537	21.2087	15.2055	37.1024
0.8	1.2	0.2434	0.2170	9.5206	15.2156	6.4195	21.9283	0.8538	6.9935	14.7708	31.9502	17.0446	44.5315
1.2	1.6	0.1455	0.1468	7.5546	13.2227	5.8378	21.0427	0.8124	6.9743	12.6885	31.7970	14.3571	41.5717
1.6	2.0	0.0828	0.0814	5.2457	10.3730	4.5988	18.3749	0.7029	6.2898	9.4502	27.6230	10.6380	35.2978
2.0	2.4	0.0364	0.0399	2.7281	5.9739	2.7400	12.5573	0.4821	4.7373	5.3163	18.7626	5.9915	23.4393
2.4	2.8	0.0120	0.0112	1.0968	2.7668	1.2710	6.6983	0.2490	2.6886	2.3244	9.8688	2.6315	12.2398
2.8	3.2	0.0034	0.0000	0.4124	1.1791	0.5407	3.2199	0.1248	1.5653	0.9459	4.9112	1.0836	6.0282
3.2	3.6	0.0014	0.0016	0.1353	0.4085	0.1953	1.3227	0.0505	0.7372	0.3305	1.9881	0.3836	2.4971
3.6	4.0	0.0003	0.0000	0.0472	0.1324	0.0736	0.5377	0.0210	0.3048	0.1217	0.7802	0.1424	0.9829
4.0	4.4	0.0002	0.0016	0.0174	0.0622	0.0256	0.1803	0.0097	0.1659	0.0445	0.3319	0.0530	0.4117
4.4	4.8	0.0001	0.0000	0.0067	0.0239	0.0100	0.0989	0.0035	0.0574	0.0169	0.1388	0.0203	0.1819
4.8	5.2	0.0000	0.0000	0.0028	0.0207	0.0046	0.0335	0.0010	0.0191	0.0066	0.0543	0.0084	0.0734
5.2	5.6	0.0000	0.0000	0.0008	0.0032	0.0018	0.0191	0.0006	0.0096	0.0026	0.0271	0.0033	0.0319
5.6	6.0	0.0000	0.0000	0.0004	0.0000	0.0009	0.0064	0.0003	0.0080	0.0014	0.0128	0.0016	0.0144
0.0	6.0	1.1771	1.0850	39.7540	68.3170	28.7810	109.8344	4.1788	37.6497	60.3478	154.1170	73.9326	217.9376

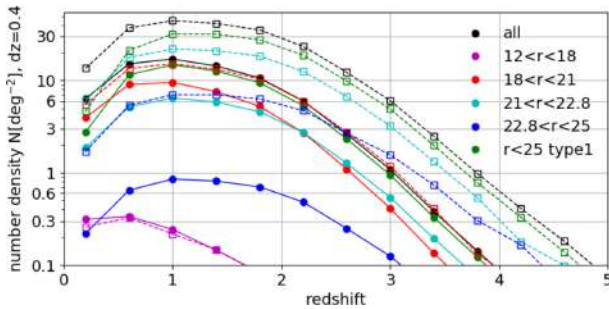


Figure 15. Predicted redshift distributions per deg^2 (intervals of $dz = 0.4$) for the fiducial eRASS8 scenario and for different magnitude cuts in the SDSS r -band: $12 < r < 18$, $18 < r < 21$, $21 < r < 22.8$, $22.8 < r < 25$, all corresponding to Table 3. The extra-Galactic distribution is represented with filled circles and the poles with empty squares. This was computed over an area of $27\,143$ (626.7) deg^2 for the extra-Galactic sky (poles), namely $|b_{\text{gal}}| > 20^\circ$ ($|b_{\text{ecl}}| > 80^\circ$).

than what is found in Myers et al. (2015) and Rodríguez-Torres et al. (2017).

Applying the Fisher formalism (Font-Ribera et al. 2014; Zhao et al. 2016), we find that the SNR3 sample (over $28k \text{ deg}^2$ i.e. the full extra-Galactic sky) will provide a BAO measurement better than 3 per cent from redshift 0.5–2 using the AGN autocorrelation and the same accuracy from redshift 2 to 3 using the Ly α forest (see Fig. 20), in good agreement with Hütsi et al. (2014b). Dedicated BAO/clustering experiments such as DESI, 4MOST and Euclid will of course provide better formal constraints, but eROSITA has the advantage that it will be based on a simple X-ray target selection. This may reduce systematics in the measurements of the clustering on large scales, or at minimum provide a cross-check to be compared to other clustering measurements.

4.5 Clustering on small scales: the constraining power of eROSITA

Exploring clustering as a function of various AGN parameter receives a large interest in the community. Retana-Montenegro & Röttgering (2017) studied the variation in the AGN clustering signal as a function of radio-loudness; Shen et al. (2013); He et al. (2018) versus optical luminosity; Alleinato et al. (2014); Jiang et al. (2016); Krumpel et al. (2018) versus AGN type; DiPompeo, Hickox & Myers (2016) versus obscuration; Krumpel, Miyaji & Coil (2010); Koutoulidis et al. (2013); Mountrichas et al. (2016) versus X-ray luminosity; Mendez et al. (2016) versus the wavelength AGNs were selected in IR or X-ray or radio; Shen et al. (2009); Komiya et al. (2013); Krumpel et al. (2015) as a function of black hole mass, Krumpel et al. (2015) versus Eddington ratio. Among others, DiPompeo et al. (2016) found a hint of a discrepancy between the clustering of obscured versus unobscured AGNs at redshift 1 (one large redshift bin $0 < z < 2$ over $\sim 3000 \text{ deg}^2$). They find obscured AGNs to have a bias of $b = 2.13 \pm 0.14$ and unobscured of $b = 1.82 \pm 0.11$. Uncertainties (~ 0.1 – 0.15) are large due to the use of angular clustering integrated over a large redshift range.

We predict the eROSITA performance relative to the DiPompeo et al. (2016) measurements. We split the AGN population into optically obscured and optically unobscured sources in the eRASS8 mock catalogue over half of the extra-Galactic sky and in redshift bins of width $dz = 0.4$ (considering only sources with $r < 23.5$ i.e. those that will have reliable spectroscopic redshifts). We count pairs with separations between 3.4 and 46.7 Mpc in 15 logarithmic bins of separation. Assuming an uncertainty dominated by Poisson noise, we evaluate the fractional difference in the correlation function that the upcoming eRASS8 AGN sample will be sensitive to, as a function of scale and redshift. We find that at low redshift, where both obscured and unobscured AGNs are present in large numbers, the eRASS8 data will be sensitive to percent level variations of the clustering amplitude at separations greater than 20 Mpc (see Fig. 21). This implies that we should easily be able to measure the suggested deviation in DiPompeo et al. (2016) at high significance.

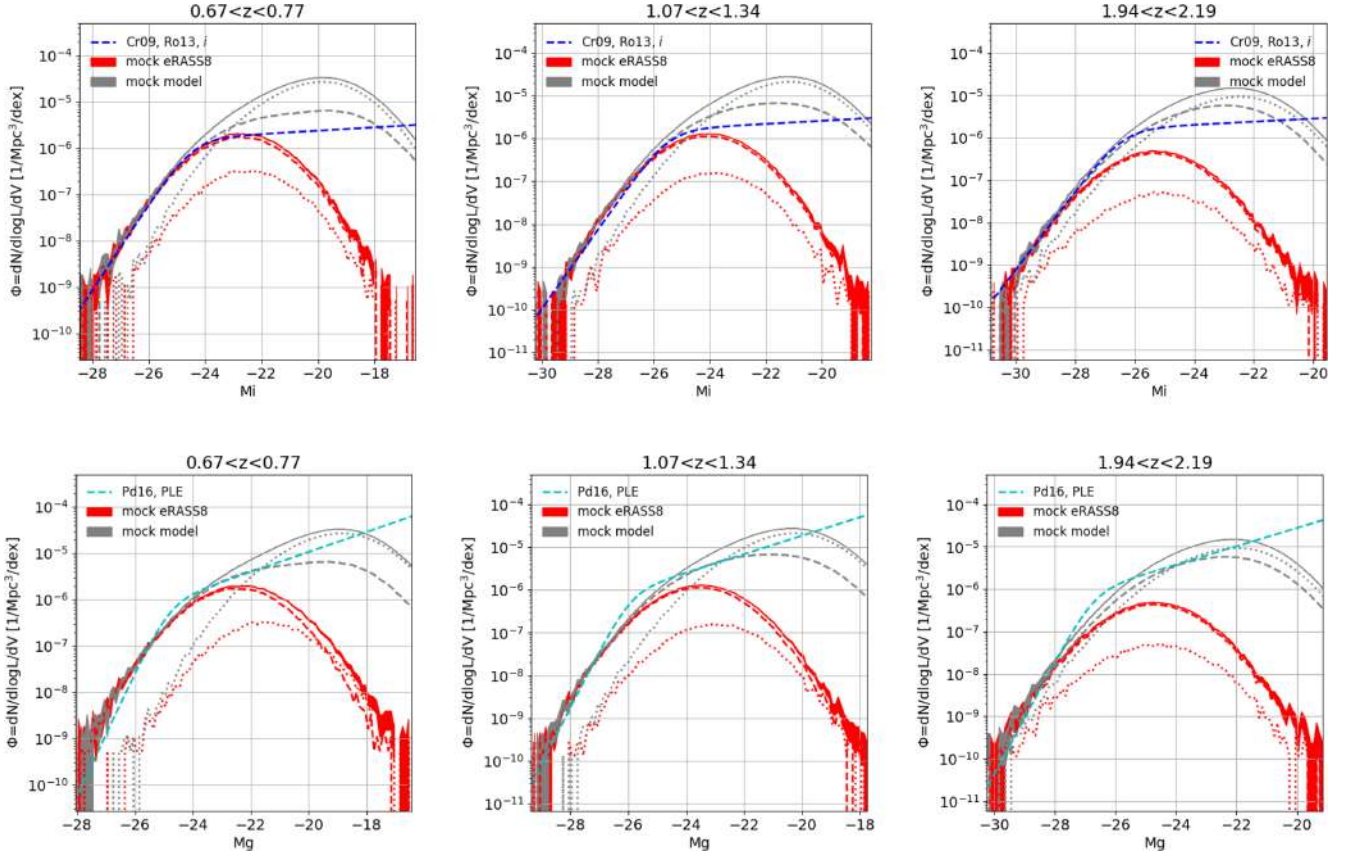


Figure 16. Observed-frame optical luminosity functions of AGNs (colour-shaded area) subdivided into optical type 1 AGNs (dashed) and type 2 and 3 AGNs (dotted). The full model, down to the resolution of the simulation is shown in grey and the eRASS8 mock in red. It is compared to the models based on observations from Croton (2009); Ross et al. (2013) (Cr09, Ro13) for the i -band (top row) and to the Caditz (2017); Palanque-Delabrouille et al. (2016) model for the g -band (bottom row). The agreement at the bright end is excellent.

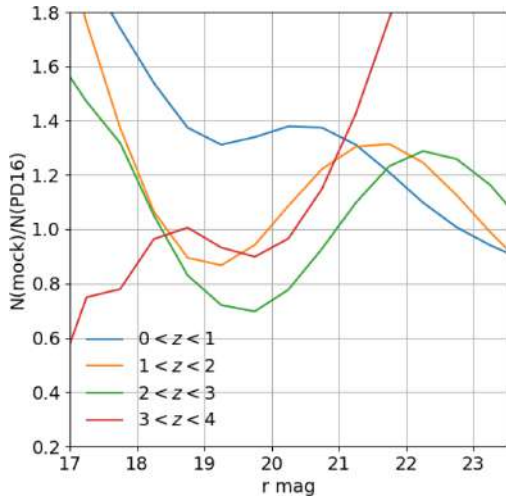


Figure 17. Ratio of the number of optical type-1 AGNs predicted as a function of r -band magnitude in wide redshift bins by the mock and by the Caditz (2017); Palanque-Delabrouille et al. (2016) models (table 7). In the range $0 < z < 3$, $19 < r < 23$, the mock is within ± 40 per cent of the models. Outside of this range, the discrepancy can reach a factor of 2.

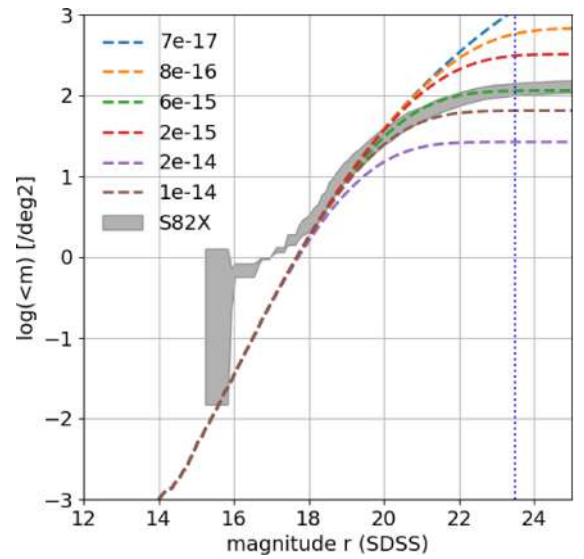


Figure 18. $\log N$ – $\log R$ relation obtained from the mock catalogue compared to the relation obtained with the Stripe82X catalogue from Ananna et al. (2017, grey shaded area). For the mock, we show many lines corresponding to different X-ray flux limits $F_X^{0.5-2\text{keV}} > 10^{-14}$, 2×10^{-14} , 2×10^{-15} , 6×10^{-15} , 8×10^{-16} , $7 \times 10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1}$.

Table 4. Number of fibre-hours needed to follow-up spectroscopically all eROSITA AGNs in the optical band. D : diameter of the telescope in meters. Magnitude: magnitude selection range. Fibre-hours: total number of fibre-hours to cover the extra-Galactic area, or the poles and number density of fibre-hours.

D (m)	Magnitude		Fibre-hours	
	Min	Max	Total	Density (deg ⁻²)
eRASS8 Xgal				
1 m	12.0	18.0	4470.67	0.17
2 m	18.0	21.0	566 463.87	21.34
4 m	21.0	22.8	1094 538.60	41.24
8 m	22.8	25.0	406 647.20	15.32
eRASS8 poles				
1 m	12.0	18.5	374.40	0.62
2 m	18.5	21.5	67 921.70	113.2
4 m	21.5	23.0	156 087.72	260.15
8 m	23.0	26.0	112 546.52	187.58

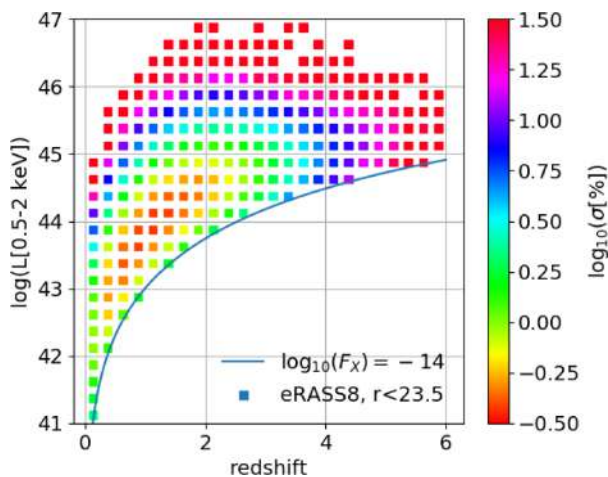


Figure 19. Fractional error [$\log_{10}(N^{-0.5})$] on number density measurements in bins of redshift and soft X-ray luminosity (0.5–2 keV band) for soft fluxes brighter than 10^{-14} erg cm⁻² s⁻¹ in the 0.5–2 keV band (the typical eROSITA flux limit).

We will furthermore be able to measure accurate differences as a function of redshift and make evolutionary statements. The current mock does not have a satellite model, we thus cannot predict the clustering at scales smaller than 5 Mpc. At higher redshift, the density of AGNs drops so, for example, we will only be able distinguish deviations at the 10 percent level from one another at redshift 2.

5 DISCUSSION AND CONCLUSIONS

5.1 Enhancing the eROSITA science with end-to-end simulations

In this section, we discuss how the mock should help to interpret the eROSITA observations. The mock catalogue is a test bench to validate algorithms either in a forward modelling approach (convolution of observational biases) or in a backward modelling fashion (deconvolution) (e.g. Ross et al. 2017; Elvin-Poole et al. 2018; Zhai et al. 2019). Here, we walk through the forward model approach that we envision using to create the eROSITA AGN large-scale structure catalogue.

The first step is to simulate the photon events corresponding to the AGN catalogue. Then feed event files in the general eROSITA pipeline to create an ‘observed’ AGN catalogue. This will contain positions, fluxes, and detection significance. The direct comparison of the input mock and the observed catalogue will allow further refinement of the sensitivity curves obtained by Clerc et al. (2018b). In particular, we will measure and tabulate the sensitivity curves and their variations with right ascension, declination, exposure time, and detection significance. Additionally, the comparison of the observed catalogue with the input catalogue will enable us to measure position uncertainties relative to true positions, and biases in the recovered fluxes as a function of AGN type, redshift, and luminosity. It will also fully justify the detection probability thresholds to be used as a function of the completeness and purity desired in any given sample. We will study how these quantities converge as a function of the total exposure time and the number of passes (i.e. from eRASS1 to eRASS8). The mock catalogue will also allow quantification of the impact of source confusions on the construction of the catalogue at the poles, where the exposure is deepest, accounting properly for clustering of the sources.

A second important application is to test the optical infrared counterpart finding algorithm (Salvato et al. 2018b) and the photometric redshift estimates (Salvato et al. 2011). For the first step, having an accurate X-ray position uncertainty is very important, but the counterpart identification process can also be tested explicitly by adding simulated optical counterparts of mock X-ray sources to real optical identification catalogues. This will allow an assessment the convergence, purity, and completeness of the counterpart finding algorithm as a function of

- (i) the X-ray parameters determined in the previous step;
- (ii) depth of the optical and IR surveys: the depth of the current realization is tuned for the combination of HSC, VHS, and WISE;
- (iii) the environment: density of galaxies, type of galaxies.

Once the counterparts to the synthetic sources are identified, their photometric redshifts can be estimated using the standard tools and compared with the known redshift from the mock catalogue. This will give a robust estimate of the photometric redshift error and the fraction of catastrophic outliers without the need for extensive spectroscopy and without limitation in terms of object numbers. Current estimates suggest that the fraction of catastrophic outliers may be as large as 15 percent with the assumed bands. The mock catalogue will allow the creation of sub samples with the most reliable photometric redshifts and/or refinement of the methods. It should also be possible to test whether studies of clustering with photometric redshifts can give reliable results.

A further application is the validation of the measured (photometric or spectroscopic) redshift distributions as a function of the position on the sky and the relative depths of the multiwavelength surveys used. This will permit the quantifications of the completeness achieved as a function of redshift and position on the sky. Because the mock is able reproduce the input luminosity function, we should derive completeness weights with good accuracy. These then feed into the measurements of the luminosity function from the real data, allowing the a measurement of the luminosity function exploiting the full statistical power of the data, uninhibited by completeness effects.

We will also use the mock to validate angular clustering measurement as a function of X-ray luminosity thresholds e.g. using volume-limited AGN samples, which are very sparse in current surveys. The discovery space around these upcoming clustering measurements is very large. On top of the very precise large-scale halo bias

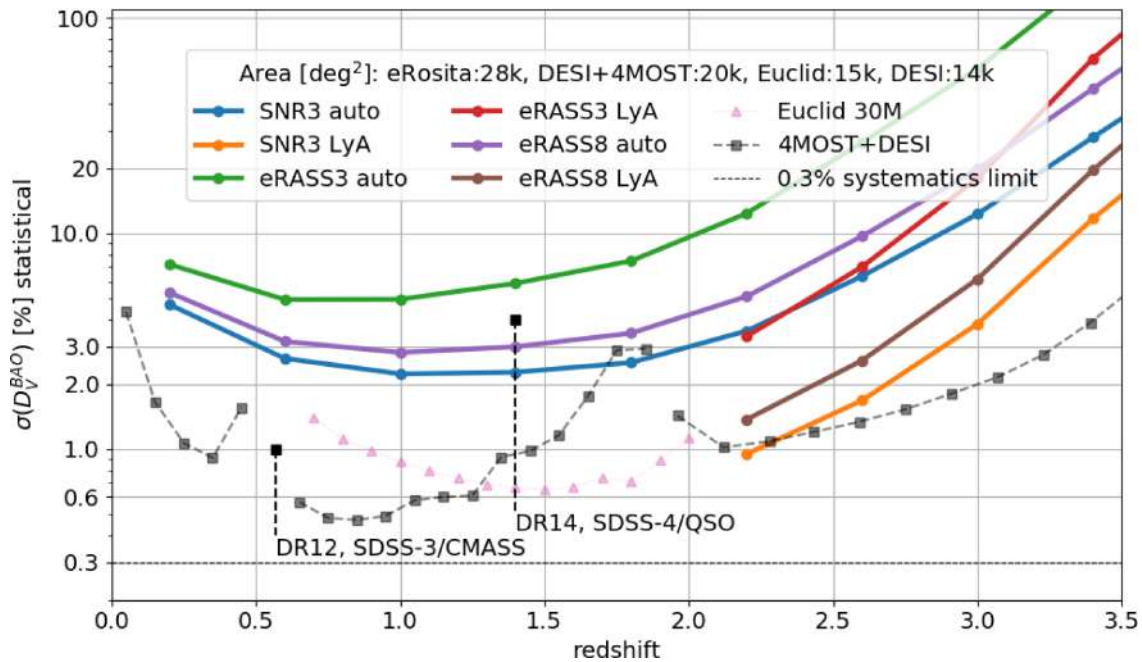


Figure 20. Expected statistical uncertainty on the angle averaged BAO scale for four surveys: eROSITA, DESI, Euclid, and DESI+4MOST. The areas covered in square degrees area: eROSITA: 28k, DESI: 14k, 4MOST: 6k (DESI + 4MOST: 20k), Euclid: 15k (30e6 galaxies scenario). For eROSITA, we consider all eRASS3, eRASS8, and SNR3 scenarios.

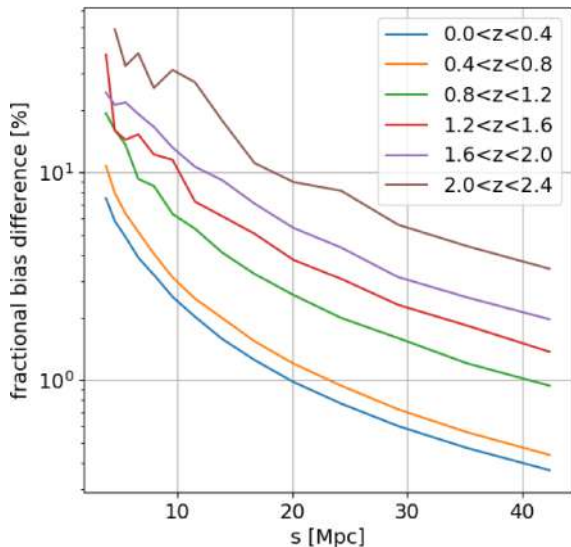


Figure 21. Sensitivity to changes in the clustering amplitude (bias) as a function of scale. Only Poisson noise is considered.

measurement we expect, cross-correlation with other galaxies at smaller scales will contain a vast amount of information and will be a powerful tool to constrain how galaxies co-evolve with AGNs (Di Matteo, Springel & Hernquist 2005; Hopkins et al. 2006, 2008; Marulli et al. 2008; Bournaud et al. 2011; Fanidakis et al. 2011, 2013; Heckman & Best 2014). To measure the clustering, we need a set of random points that are imprinted with the sensitivity map and systematic biases induced by the two steps described above. To do so we will run the same analysis as described above except that right ascension and declination will be randomly distributed on the celestial sphere. In this manner we obtain a set of random points that by construction contains all the possible systematics. The

forward modelling approach allows clustering measurements down to any angular separation while the backward modelling approach has a minimum scale below which clustering cannot be recovered. Additional milky way, stellar, dust masks, and spectroscopic follow-up selection function that add a weight to entire regions of the sky are applied a posteriori on both data and random positions.

In summary, this forward model approach guarantees an optimal control over possible systematic errors affecting the estimation of the number density as a function of sky position, redshift, and depth (X-ray or optical and IR).

5.2 On small-scale clustering

While there is clear observational evidence that supermassive black holes and galaxies evolve together, the physical reason remains unclear (Gebhardt et al. 2000; Ferrarese & Merritt 2000; Brandt & Alexander 2015). There are two main paths of co-evolution. In one, the AGN evolves jointly with its galaxy host, with its accretion rate to some extent correlated to star formation rate (Di Matteo et al. 2005; Hopkins et al. 2006, 2008; Marulli et al. 2008). In the other, AGNs would occur randomly in the galaxy population with some duty cycle (Bournaud et al. 2011; Bongiorno et al. 2016), with some kind of self-regulatory mechanism imprinting the BH–galaxy correlation. In the future, we intend to extend our model to create predictions for the different scenarios of AGN–host galaxy coevolution. We will use SAMS the galaxy population containing central and satellite galaxies (Lacey et al. 2016). Our empirical AGN X-ray model then shows us how to add X-ray AGNs to the SAM model and study AGN relative to other galaxy populations (e.g. Bonoli et al. 2009) that can subsequently be searched for and tested.

5.3 Summary and conclusion

We have described a new methodology to construct mock catalogues based on dark matter only N -body simulations for X-ray-selected AGN samples. By construction, the model reproduces accurately (± 5 per cent) the 2–10 keV luminosity functions up to redshift 6 as well as the observed AGN log N –log S function (to within ± 20 per cent). We create an empirical AGN obscuration model that jointly classifies AGNs in the optical and the X-ray. From the observed distribution of AGNs in the X-ray flux – optical r -band magnitude plane, we predict optical properties for AGNs, obtaining reasonable agreement with the most recent published optical AGN luminosity functions (Ross et al. 2013; Palanque-Delabrouille et al. 2016; Caditz 2017). Finally, we tabulate a set of X-ray spectra, UV–optical–IR SEDs and optical medium resolution spectra that represent well the AGN population present in the mock catalogue.

This is used to predict the redshift distribution of the AGN to be detected by eROSITA, showing good agreement with previous estimates (Merloni et al. 2012; Hütsi, Gilfanov & Sunyaev 2014a). We show that the XLF will be recovered to an extremely high level of accuracy, more than a factor of 10 better than current state-of-the-art estimates. Our model also predicts the large-scale three-dimensional distribution of the eROSITA AGN. The large-scale bias of halo hosting AGNs is in agreement with previous estimates (Krumpe et al. 2015, 2018; Chehade et al. 2016; Rodríguez-Torres et al. 2017). Once again the eROSITA survey will yield clustering measurements of X-ray-selected AGN with unprecedented accuracy. Finally, we predict the measurements expected of the BAO standard ruler and find it will be a very interesting data set for large-scale structure studies, complementing dedicated cosmological experiments, but with arguably a simple selection function.

The mock catalogue is a useful tool for full and accurate science exploitation of the eROSITA all-sky survey: It (1) provides accurate forecasts and estimates of the power of the eROSITA survey, (2) can validate analysis methods, and (3) estimate a wide variety of errors, biases, incompletenesses, and correction factors.

ACKNOWLEDGEMENTS

J. Comparat thanks M. Freyberg, K. Dennerl, S. Bonoli, H. Guo, J. Aird, and T. Miyaji for fruitful discussions about this project.

JB acknowledges support from the CONICYT-Chile grants Basal-CATA PFB-06/2007 & AFB-170002, FONDECYT Postdoctorados 3160439 and the Ministry of Economy, Development, and Tourism’s Millennium Science Initiative through grant IC120009, awarded to The Millennium Institute of Astrophysics, MAS. This research was supported by the DFG cluster of excellence ‘Origin and Structure of the Universe’.

GY would like to thank MINECO/FEDER for financial support under project grant AYA2015-63810-P.

MK acknowledges support from DLR grant 50OR1802.

The CosmoSim data base is a service of the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark data base was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064.

REFERENCES

Abazajian K. N. et al., 2009, *ApJS*, 182, 543
 Abolfathi B. et al., 2018, *ApJS*, 235, 42
 Aihara H. et al., 2018, *PASJ*, 70, S8
 Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, *MNRAS*, 451, 1892

Aird J., Coil A. L., Georgakakis A., 2018, *MNRAS*, 474, 1225
 Allevalo V. et al., 2014, *ApJ*, 796, 4
 Ananna T. T. et al., 2017, *ApJ*, 850, 66
 Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, *MNRAS*, 426, 2046
 Bacon R. et al., 2015, *A&A*, 575, A75
 Behroozi P., Wechsler R., Wu H.-Y., 2013, *ApJ*, 762, 109
 Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
 Biffi V., Dolag K., Merloni A., 2018, *MNRAS*, 481, 2213
 Bolton A. S. et al., 2012, *AJ*, 144, 144
 Bongiorno A. et al., 2016, *A&A*, 588, A78
 Bonoli S., Marulli F., Springel V., White S. D. M., Branchini E., Moscardini L., 2009, *MNRAS*, 396, 423
 Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, *ApJ*, 741, L33
 Brandt W. N., Alexander D. M., 2015, *A&AR*, 23, 1
 Buchner J. et al., 2015, *ApJ*, 802, 89
 Buchner J., Bauer F. E., 2017, *MNRAS*, 465, 4348
 Buchner J., Nandra K., Bauer F. E., Brightman M., Nikutta R., 2018, in 42nd COSPAR Scientific Assembly. E1.6–60–18
 Caditz D. M., 2017, *A&A*, 608, A64
 Caplar N., Lilly S. J., Trakhtenbrot B., 2015, *ApJ*, 811, 148
 Caplar N., Lilly S. J., Trakhtenbrot B., 2018, *ApJ*, 867, 148
 Chehade B. et al., 2016, *MNRAS*, 459, 1179
 Ciesla L. et al., 2015, *A&A*, 576, A10
 Clerc N. et al., 2018a, *A&A*, 617, A92
 Clerc N. et al., 2018b, *A&A*, 617, A92
 Comastri A. et al., 2002, *ApJ*, 571, 771
 Comparat J. et al., 2013, *MNRAS*, 428, 1498
 Comparat J. et al., 2016, *A&A*, 592, A121
 Comparat J., Prada F., Yepes G., Klypin A., 2017, *MNRAS*, 469, 4157
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
 Croton D. J., 2009, *MNRAS*, 394, 1109
 da Cunha E. et al., 2017, *PASA*, 34, e047
 Davidzon I. et al., 2017, *A&A*, 605, A70
 Dawson K. S. et al., 2013, *AJ*, 145, 10
 Dawson K. S. et al., 2016, *AJ*, 151, 44
 de Jong R. S. et al., 2012, Proc. SPIE, 8446, 84460T
 DESI Collaboration et al., 2016, preprint (arXiv:1611.00036)
 Dey A. et al., 2019, *AJ*, 157, 168
 Di Matteo T., Springel V., Hernquist L., 2005, *Nature*, 433, 604
 DiPompeo M. A., Hickox R. C., Myers A. D., 2016, *MNRAS*, 456, 924
 Dwelly T. et al., 2017, *MNRAS*, 469, 1065
 Dwelly T., Merloni A., Walcher J. C., Clerc N., Gueguen A., Boller T., de Jong R. S., Chiappini C., 2016, Observatory Operations: Strategies, Processes, and Systems VI. SPIE, p. 99101Q
 Edge A. et al., 2013, *The Messenger*, 154, 32
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
 Elvin-Poole J. et al., 2018, *Phys. Rev. D*, 98, 042006
 Fanidakis N. et al., 2013, *MNRAS*, 435, 679
 Fanidakis N., Baugh C. M., Benson A. J., Bower R. G., Cole S., Done C., Frenk C. S., 2011, *MNRAS*, 410, 53
 Favole G. et al., 2016, *MNRAS*, 461, 3421
 Ferrarese L., Merritt D., 2000, *ApJ*, 539, L9
 Fitzpatrick E. L., 1999, *PASP*, 111, 63
 Fliri J., Trujillo I., 2016, *MNRAS*, 456, 1359
 Font-Ribera A., McDonald P., Mostek N., Reid B. A., Seo H.-J., Slosar A., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 023
 Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
 Gebhardt K. et al., 2000, *ApJ*, 539, L13
 Georgakakis A., Nandra K., Laird E. S., Aird J., Trichas M., 2008, *MNRAS*, 388, 1205
 Georgakakis A., Aird J., Schulze A., Dwelly T., Salvato M., Nandra K., Merloni A., Schneider D. P., 2017, *MNRAS*, 471, 1976
 Georgakakis A., Comparat J., Merloni A., Ciesla L., Aird J., Finoguenov A., 2018, *MNRAS*, 3272 (G18)
 Guo H. et al., 2019, *ApJ*, 871, 147

- Guo Q., White S., Li C., Boylan-Kolchin M., 2010, *MNRAS*, 404, 1111
- He W. et al., 2018, *PASJ*, 70, S33
- Heckman T. M., Best P. N., 2014, *ARA&A*, 52, 589
- Heitmann K. et al., 2015, *ApJS*, 219, 34
- Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, *MNRAS*, 367, 454
- HI4PI Collaboration et al., 2016, *A&A*, 594, A116
- Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, *MNRAS*, 442, 2304
- Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, *ApJS*, 163, 1
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, *ApJS*, 175, 356
- Hsu L.-T. et al., 2014, *ApJ*, 796, 60
- Hütsi G., Gilfanov M., Sunyaev R., 2014a, *A&A*, 561, A58
- Hütsi G., Gilfanov M., Kolodzig A., Sunyaev R., 2014b, *A&A*, 572, A28
- Ilbert O. et al., 2013, *A&A*, 556, A55
- Ishiyama T., Enoki M., Kobayashi M. A. R., Makiya R., Nagashima M., Oogi T., 2015, *PASJ*, 67, 61
- Jiang N., Wang H., Mo H., Dong X.-B., Wang T., Zhou H., 2016, *ApJ*, 832, 111
- Kawanomoto S. et al., 2018, *PASJ*, 70, 66
- Klypin A., Prada F., 2018, *MNRAS*, 478, 4602
- Klypin A., Prada F., Yepes G., Heß S., Gottlöber S., 2015, *MNRAS*, 447, 3693
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Kollmeier J. A. et al., 2017, preprint (arXiv e-prints)
- Kolodzig A., Gilfanov M., Sunyaev R., Sazonov S., Brusa M., 2013, *A&A*, 558, A89
- Komiya Y., Shirasaki Y., Ohishi M., Mizumoto Y., 2013, *ApJ*, 775, 43
- Koulouridis E. et al., 2018, *A&A*, 620, A4
- Koutoulidis L., Plionis M., Georgantopoulos I., Fanidakis N., 2013, *MNRAS*, 428, 1382
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Krumpe M., Miyaji T., Coil A. L., 2010, *ApJ*, 713, 558
- Krumpe M., Miyaji T., Husemann B., Fanidakis N., Coil A. L., Aceves H., 2015, *ApJ*, 815, 21
- Krumpe M., Miyaji T., Coil A. L., Aceves H., 2018, *MNRAS*, 474, 1773
- Lacey C. G. et al., 2016, *MNRAS*, 462, 3854
- Marchesi S. et al., 2016, *ApJ*, 817, 34
- Martin D. C. et al., 2005, *ApJ*, 619, L1
- Marulli F., Bonoli S., Branchini E., Moscardini L., Springel V., 2008, *MNRAS*, 385, 1846
- Mateos S. et al., 2008, *A&A*, 492, 51
- Mayer L., Bonoli S., 2019, *Rep. Progr. Phys.*, 82, 016901
- McAlpine S., Bower R. G., Harrison C. M., Crain R. A., Schaller M., Schaye J., Theuns T., 2017, *MNRAS*, 468, 3395
- McConnachie A. et al., 2016, preprint (arXiv:1606.00043)
- Mendez A. J. et al., 2016, *ApJ*, 821, 55
- Menzel M.-L. et al., 2016, *MNRAS*, 457, 110
- Merloni A. et al., 2012, preprint (arXiv:1209.3114)
- Merloni A. et al., 2014, *MNRAS*, 437, 3550
- Merloni A. et al., 2019, *Messenger*, 175, 42
- Miyaji T. et al., 2015, *ApJ*, 804, 104
- Morrison R., McCammon D., 1983, *ApJ*, 270, 119
- Morrissey P. et al., 2007, *ApJS*, 173, 682
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Mountrichas G., Georgakakis A., Menzel M.-L., Fanidakis N., Merloni A., Liu Z., Salvato M., Nandra K., 2016, *MNRAS*, 457, 4195
- Myers A. D. et al., 2015, *ApJS*, 221, 27
- Padovani P. et al., 2017, *A&AR*, 25, 2
- Palanque-Delabrouille N. et al., 2016, *A&A*, 587, A41
- Pâris I. et al., 2018, *A&A*, 613, A51
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Prakash A. et al., 2016, *ApJS*, 224, 34
- Predehl P. et al., 2016, in SPIE Conf. Ser. Vol. SPIE, Bellingham, p. 99051K
- Raichoor A. et al., 2017, *MNRAS*, 471, 3955
- Reid B. et al., 2016, *MNRAS*, 455, 1553
- Retana-Montenegro E., Röttgering H. J. A., 2017, *A&A*, 600, A97
- Ricci C. et al., 2017, *Nature*, 549, 488
- Richard J. et al., 2019, *Messenger*, 175, 50
- Rodríguez-Torres S. A. et al., 2016, *MNRAS*, 460, 1173
- Rodríguez-Torres S. A. et al., 2017, *MNRAS*, 468, 728
- Ross A. J. et al., 2017, *MNRAS*, 464, 1168
- Ross N. P. et al., 2013, *ApJ*, 773, 14
- Salvato M. et al., 2011, *ApJ*, 742, 61
- Salvato M. et al., 2018b, *MNRAS*, 473, 4937
- Salvato M., Ilbert O., Hoyle B., 2018a, *Nat. Astron.*, 68
- Scodreggio M. et al., 2005, *PASP*, 117, 1284
- Seo H.-J., Eisenstein D. J., 2003, *ApJ*, 598, 720
- Seo H.-J., Eisenstein D. J., 2007, *ApJ*, 665, 14
- Shen Y. et al., 2009, *ApJ*, 697, 1656
- Shen Y. et al., 2013, *ApJ*, 778, 98
- Skillman S. W., Warren M. S., Turk M. J., Wechsler R. H., Holz D. E., Sutter P. M., 2014, preprint (arXiv:1407.2600)
- The MSE Science Team et al., 2019, preprint (arXiv:1904.04907)
- Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, *ApJ*, 742, 16
- Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, *ApJ*, 786, 104
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- Yaqoob T., 1997, *ApJ*, 479, 184
- Zhai Z. et al., 2019, *ApJ*, 874, 95
- Zhao G.-B. et al., 2016, *MNRAS*, 457, 2377
- Zhu G., 2016, preprint (arXiv:1606.07156)

APPENDIX A: TABLES

See Tables [A1–A4](#).

Table A1. Same as Table 3 for the eRASS3 scenario.

z bin		$12 < r < 18$		$18 < r < 21$		$21 < r < 22.8$		$22.8 < r < 25$		type 1 $r < 25$		All	
Min	Max	x	p	x	p	x	p	x	p	x	p	x	p
0.0	0.4	0.3075	0.2569	2.9120	4.6591	0.8473	3.5422	0.0620	0.6095	1.9963	3.5917	4.1294	9.0901
0.4	0.8	0.3156	0.3255	6.1014	11.2649	2.2519	10.0411	0.1736	1.8573	7.3422	15.9719	8.8436	23.5302
0.8	1.2	0.2215	0.2122	5.9180	12.1999	2.6236	12.1999	0.2243	2.3982	8.2504	21.7751	8.9892	27.0677
1.2	1.6	0.1268	0.1436	4.3450	10.2916	2.2321	11.2777	0.2090	2.4620	6.3860	20.3582	6.9142	24.2227
1.6	2.0	0.0711	0.0798	2.8172	7.7402	1.6417	9.6055	0.1751	2.1812	4.3386	16.6117	4.7065	19.6545
2.0	2.4	0.0309	0.0399	1.3617	4.2028	0.9154	6.1271	0.1093	1.4759	2.2114	10.1767	2.4182	11.8744
2.4	2.8	0.0098	0.0112	0.5110	1.8684	0.3786	2.9152	0.0506	0.7818	0.8590	4.7692	0.9503	5.5894
2.8	3.2	0.0029	0.0000	0.1834	0.7068	0.1511	1.3563	0.0242	0.4037	0.3236	2.1604	0.3621	2.4827
3.2	3.6	0.0010	0.0016	0.0569	0.2409	0.0527	0.4867	0.0096	0.1915	0.1054	0.7739	0.1204	0.9270
3.6	4.0	0.0003	0.0000	0.0193	0.0654	0.0188	0.1628	0.0036	0.0750	0.0363	0.2537	0.0420	0.3032
4.0	4.4	0.0001	0.0016	0.0067	0.0367	0.0064	0.0702	0.0014	0.0271	0.0122	0.1149	0.0146	0.1356
4.4	4.8	0.0001	0.0000	0.0025	0.0160	0.0028	0.0335	0.0006	0.0080	0.0048	0.0431	0.0059	0.0574
4.8	5.2	0.0000	0.0000	0.0010	0.0112	0.0011	0.0112	0.0001	0.0016	0.0017	0.0176	0.0022	0.0239
5.2	5.6	0.0000	0.0000	0.0003	0.0016	0.0004	0.0048	0.0001	0.0048	0.0006	0.0080	0.0009	0.0112
5.6	6.0	0.0000	0.0000	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0002	0.0000	0.0003	0.0000
0.0	6.0	1.0875	1.0722	24.2363	53.3057	11.1241	57.8340	1.0435	12.4776	31.8688	96.6261	37.4997	124.9702

Table A2. Same as Table A1 for the SNR3 scenario.

z bin		$12 < r < 18$		$18 < r < 21$		$21 < r < 22.8$		$22.8 < r < 25$		Type 1 $r < 25$		All	
Min	Max	x	p	x	p	x	p	x	p	x	p	x	p
0.0	0.4	0.3166	0.2585	4.4496	5.6899	2.6049	7.4881	0.3889	2.6966	3.1866	5.2399	7.7644	16.2304
0.4	0.8	0.3417	0.3271	10.4147	14.3077	7.3314	22.2442	1.1525	8.4950	13.8005	23.7393	19.2528	45.6548
0.8	1.2	0.2490	0.2170	11.3186	16.4027	9.2570	27.7235	1.5187	10.9857	18.6644	37.4024	22.3592	55.6719
1.2	1.6	0.1496	0.1468	9.2547	14.3460	8.6079	26.7023	1.4894	11.1484	16.7610	38.1826	19.5168	52.6786
1.6	2.0	0.0861	0.0814	6.6144	11.3766	6.9738	23.4904	1.2873	10.0235	12.9902	33.8202	14.9769	45.2974
2.0	2.4	0.0377	0.0399	3.5655	6.7111	4.3219	16.4378	0.9012	7.6381	7.6889	23.8079	8.8379	31.0822
2.4	2.8	0.0128	0.0112	1.4917	3.1752	2.0872	9.1252	0.4947	4.5474	3.5570	13.3041	4.0930	17.0202
2.8	3.2	0.0036	0.0000	0.5752	1.3802	0.9150	4.5203	0.2495	2.6918	1.5036	6.8132	1.7479	8.7072
3.2	3.6	0.0015	0.0016	0.1931	0.4898	0.3479	1.9977	0.1090	1.4025	0.5590	3.0572	0.6538	3.9491
3.6	4.0	0.0003	0.0000	0.0692	0.1755	0.1300	0.8153	0.0461	0.6159	0.2087	1.2749	0.2465	1.6387
4.0	4.4	0.0002	0.0016	0.0260	0.0830	0.0494	0.3143	0.0197	0.2856	0.0792	0.5409	0.0956	0.6909
4.4	4.8	0.0001	0.0000	0.0097	0.0335	0.0194	0.1532	0.0078	0.1340	0.0308	0.2441	0.0371	0.3271
4.8	5.2	0.0000	0.0000	0.0037	0.0223	0.0084	0.0670	0.0032	0.0543	0.0126	0.1069	0.0153	0.1436
5.2	5.6	0.0000	0.0000	0.0014	0.0032	0.0036	0.0319	0.0013	0.0176	0.0051	0.0399	0.0063	0.0543
5.6	6.0	0.0000	0.0000	0.0006	0.0000	0.0017	0.0128	0.0011	0.0160	0.0027	0.0223	0.0033	0.0287
0.0	6.0	1.1991	1.0850	47.9880	74.1968	42.6595	141.1241	7.6703	60.7523	79.0504	187.5958	99.6068	279.1750

Table A3. Same as Table 4 for the eRASS3 scenario.

D (m)	Magnitude		Fibre-hours	
	(min)	(max)	Total	Density (deg ⁻²)
eRASS3 Xgal				
1 m	12.0	18.0	4070	0.15
2 m	18.0	21.0	312 051	11.76
4 m	21.0	22.8	372 991	14.05
8 m	22.8	25.0	88 220	3.32
eRASS3 poles				
1 m	12.0	18.5	367	0.61
2 m	18.5	21.5	45 674	76.12
4 m	21.5	23.0	70 477	117.46
8 m	23.0	26.0	35 201	58.67

Table A4. Same as Table 4 for the SNR3 scenario.

D (m)	Magnitude		Fibre-hours	
	Min	Max	Total	Density (deg ⁻²)
SNR3 Xgal				
1 m	12.0	18.0	4569	0.17
2 m	18.0	21.0	711 221	26.80
4 m	21.0	22.8	1717 619	64.71
8 m	22.8	25.0	789 628	29.75
SNR3 poles				
1 m	12.0	18.5	375	0.63
2 m	18.5	21.5	78 179	130.30
4 m	21.5	23.0	215 965	359.94
8 m	23.0	26.0	187 819	313.03

APPENDIX B: FIGURES

See Figs B1–B7.

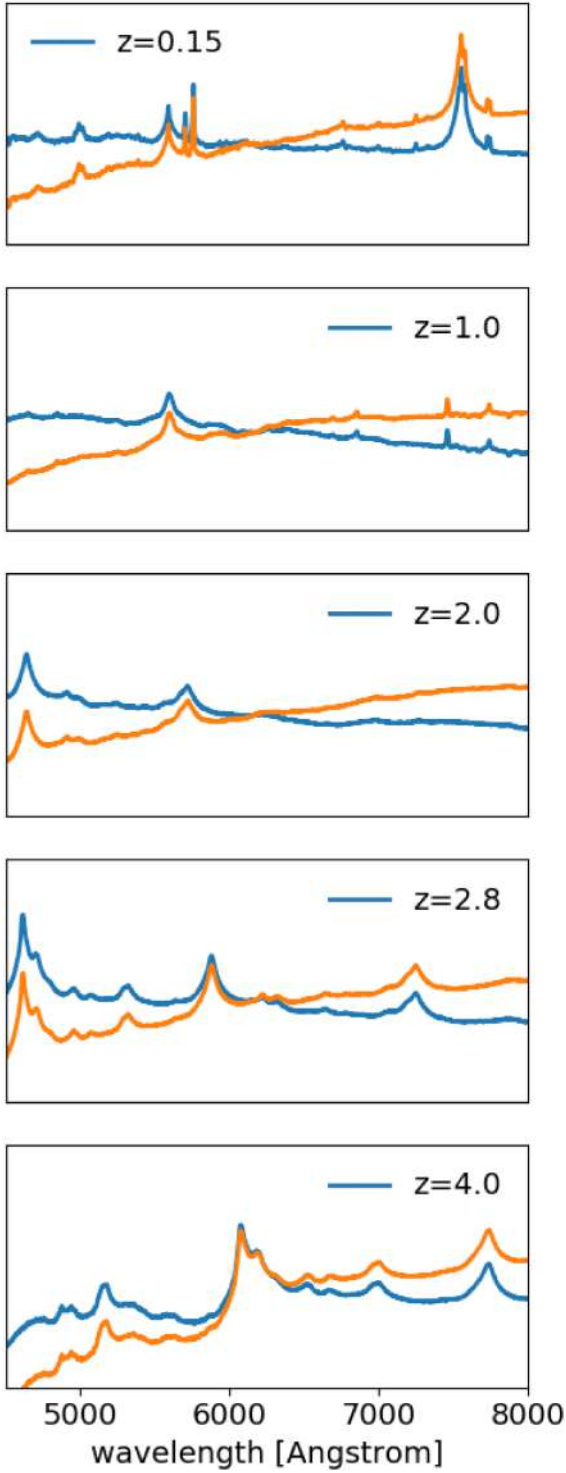


Figure B1. Flux (arbitrary units) versus wavelength of the type-1 spectral templates (blue). The orange line shows the same spectrum reddened with $E(B - V) = 1$. Both are normalized to an SDSS r band of 16.5.

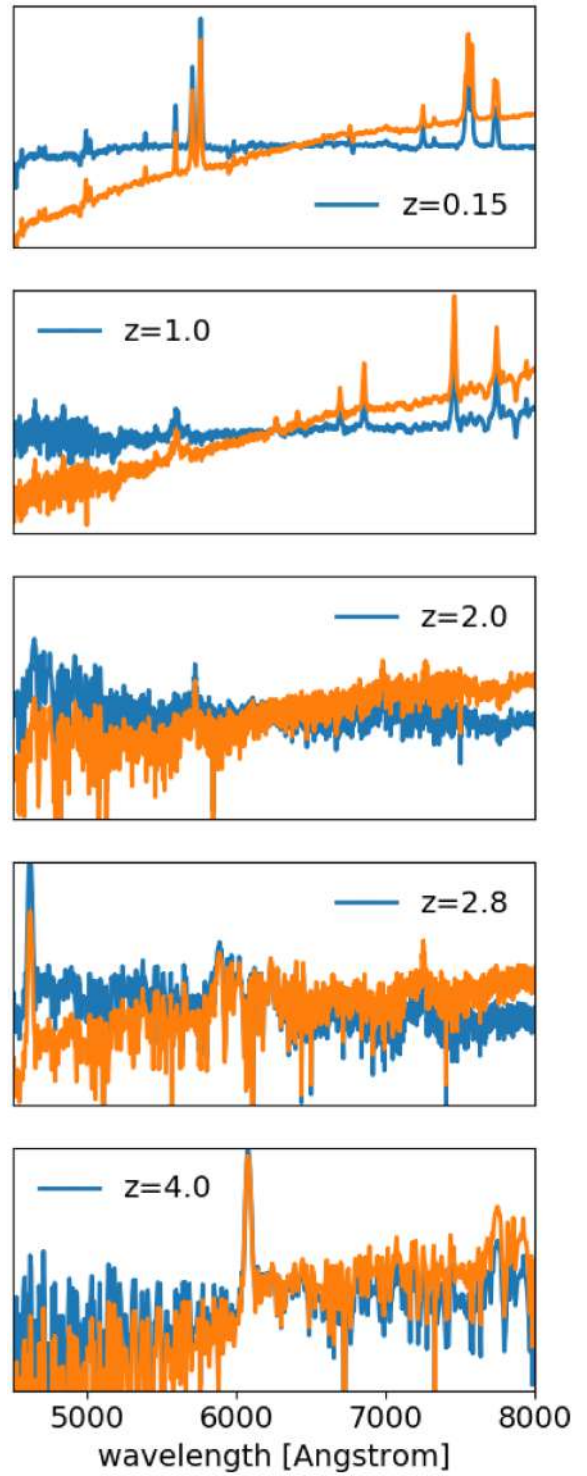


Figure B2. Same as Fig. B1 for the type-2 spectral templates.

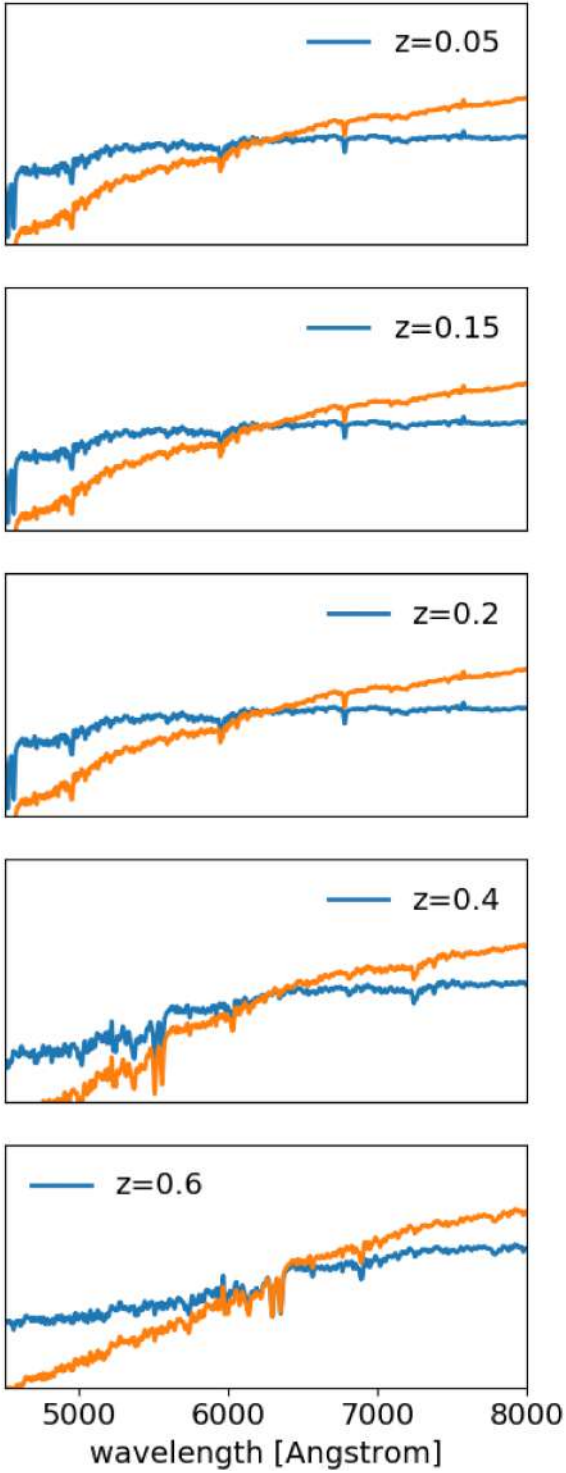


Figure B3. Same as Fig. B1 for the red elliptical galaxies, used also for the type-spectral templates.

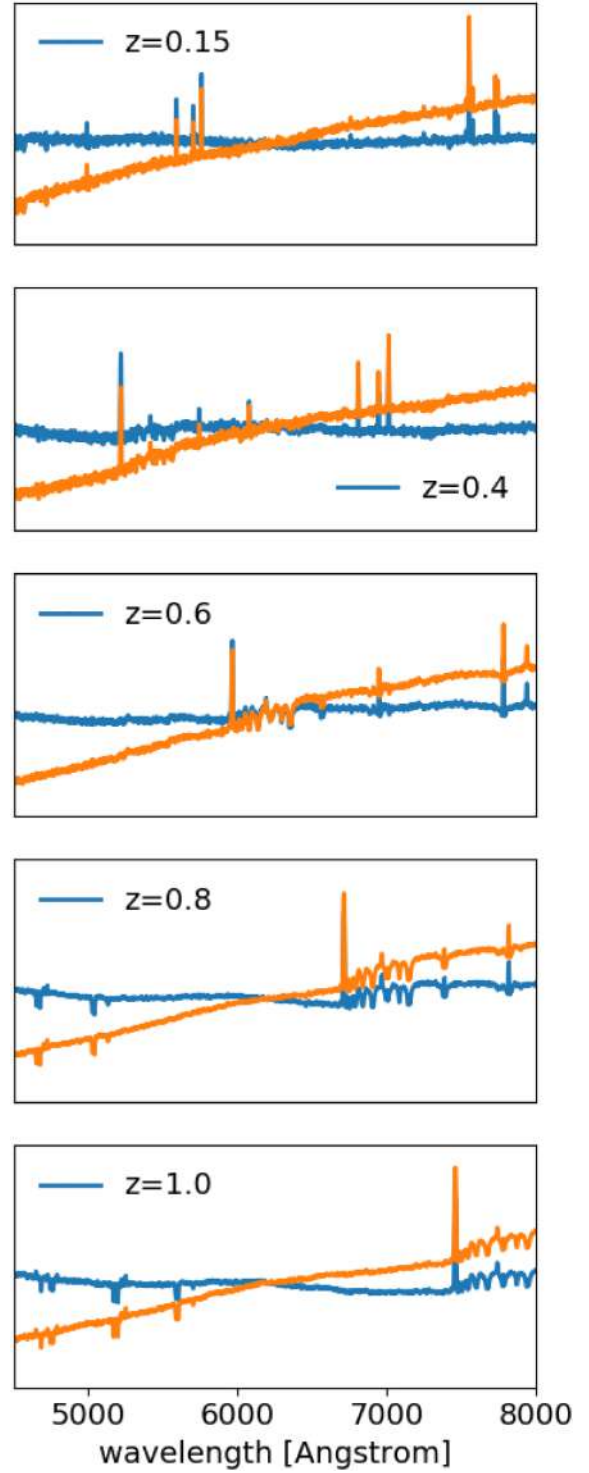


Figure B4. Same as Fig. B1 for the star-forming galaxy spectral templates. These are not used in the current mock.

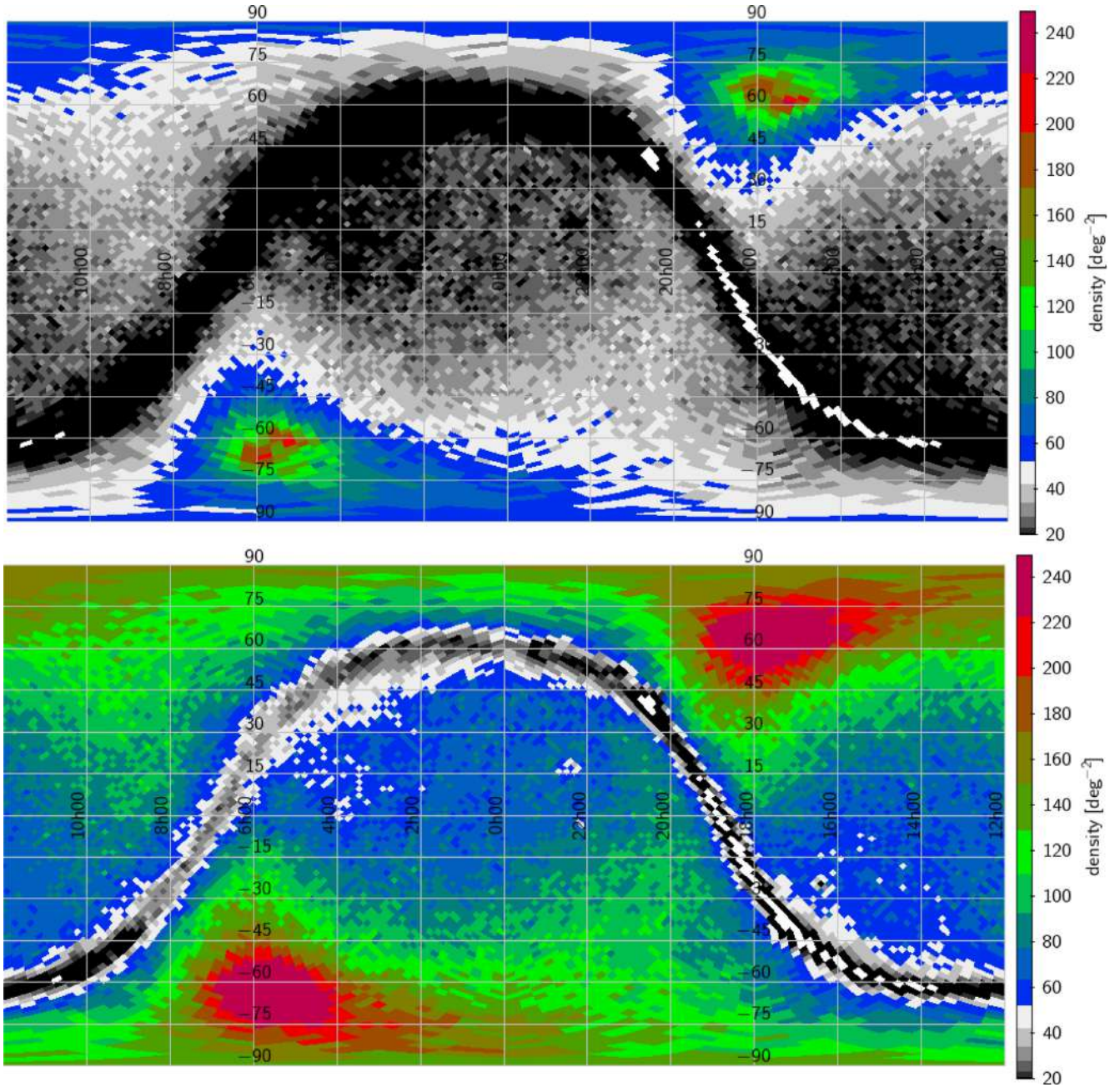


Figure B5. Same as Fig. 14. Top panel eRASS3, bottom SNR3.

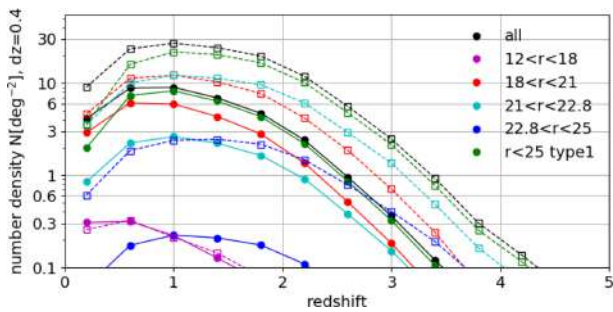


Figure B6. Same as Fig. 15 for the eRASS3 scenario. It corresponds to Table A1.

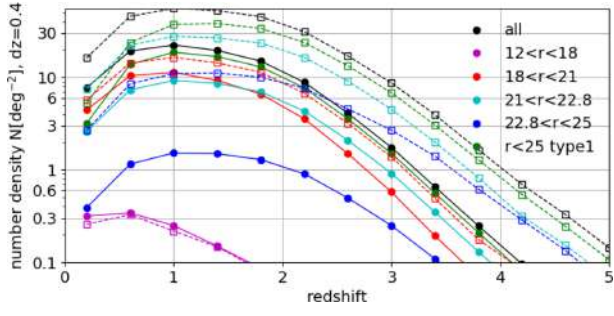


Figure B7. Same as Fig. 15 for the SNR3 scenario. It corresponds to Table A2.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.