



**HAL**  
open science

## Synthetic simulations of the extragalactic sky seen by eROSITA

N. Clerc, M.E. Ramos-Ceja, J. Ridl, G. Lamer, H. Brunner, F. Hofmann, J. Comparat, F. Pacaud, F. Käfer, T.H. Reiprich, et al.

► **To cite this version:**

N. Clerc, M.E. Ramos-Ceja, J. Ridl, G. Lamer, H. Brunner, et al.. Synthetic simulations of the extragalactic sky seen by eROSITA: I. Pre-launch selection functions from Monte-Carlo simulations. *Astronomy and Astrophysics - A&A*, 2018, 617, pp.A92. 10.1051/0004-6361/201732119 . hal-01846773

**HAL Id: hal-01846773**

**<https://hal.science/hal-01846773>**

Submitted on 11 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Synthetic simulations of the extragalactic sky seen by eROSITA

## I. Pre-launch selection functions from Monte-Carlo simulations

N. Clerc<sup>1,3</sup>, M. E. Ramos-Ceja<sup>2</sup>, J. Ridl<sup>1</sup>, G. Lamer<sup>4</sup>, H. Brunner<sup>1</sup>, F. Hofmann<sup>1</sup>, J. Comparat<sup>1</sup>, F. Pacaud<sup>2</sup>, F. Käfer<sup>1</sup>, T. H. Reiprich<sup>2</sup>, A. Merloni<sup>1</sup>, C. Schmid<sup>5</sup>, T. Brand<sup>5</sup>, J. Wilms<sup>5</sup>, P. Friedrich<sup>1</sup>, A. Finoguenov<sup>1,6</sup>, T. Dauser<sup>5</sup>, and I. Kreykenbohm<sup>5</sup>

<sup>1</sup> Max-Planck Institut für extraterrestrische Physik, Postfach 1312, 85741 Garching bei München, Germany

<sup>2</sup> Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

<sup>3</sup> IRAP, Université de Toulouse, CNRS, UPS, CNES, Toulouse, France  
e-mail: nicolas.clerc@irap.omp.eu

<sup>4</sup> Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany

<sup>5</sup> Dr. Karl Remeis-Observatory and ECAP, Sternwartstr. 7, 96049 Bamberg, Germany

<sup>6</sup> Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2a, 00014 Helsinki, Finland

Received 17 October 2017 / Accepted 21 June 2018

### ABSTRACT

**Context.** Studies of galaxy clusters provide stringent constraints on models of structure formation. Provided that selection effects are under control, large X-ray surveys are well suited to derive cosmological parameters, in particular those governing the dark energy equation of state.

**Aims.** We forecast the capabilities of the all-sky eROSITA (extended ROentgen Survey with an Imaging Telescope Array) survey to be achieved by the early 2020s. We bring special attention to modelling the entire chain from photon emission to source detection and cataloguing.

**Methods.** The selection function of galaxy clusters for the upcoming eROSITA mission is investigated by means of extensive and dedicated Monte-Carlo simulations. Employing a combination of accurate instrument characterisation and a state-of-the-art source detection technique, we determine a cluster detection efficiency based on the cluster fluxes and sizes.

**Results.** Using this eROSITA cluster selection function, we find that eROSITA will detect a total of approximately  $10^5$  clusters in the extra-galactic sky. This number of clusters will allow eROSITA to put stringent constraints on cosmological models. We show that incomplete assumptions on selection effects, such as neglecting the distribution of cluster sizes, induce a bias in the derived value of cosmological parameters.

**Conclusions.** Synthetic simulations of the eROSITA sky capture the essential characteristics impacting the next-generation galaxy cluster surveys and they highlight parameters requiring tight monitoring in order to avoid biases in cosmological analyses.

**Key words.** cosmological parameters – surveys – large-scale structure of Universe – X-rays: galaxies: clusters – cosmology: observations – methods: data analysis

## 1. Introduction

Clusters of galaxies are the most massive matter halos. They formed last in the history of the Universe by a hierarchical growth of structures in the Hubble expansion flow. Their presence, observed space density, and mass distributions confirm the standard cosmological model (e.g. Hasselfield et al. 2013; Mantz et al. 2014; Planck Collaboration XXIV 2016; de Haan et al. 2016), making galaxy clusters powerful probes of cosmological parameters, such as the dark energy content and its equation of state (e.g. Vikhlinin et al. 2009); see also Allen et al. (2011) for a review. The identification and study of the different components of galaxy clusters (dark matter halo, intra-cluster medium, galaxies, and relativistic particles) require the use of several different observational techniques. Among such techniques, X-ray observations stand out, since clusters of galaxies are the most luminous extended sources in the extra-galactic X-ray sky, and therefore are easily detectable in large surveys. The importance of galaxy clusters in a cosmological context has been realized since the pioneering surveys undertaken with

the *Einstein* observatory (e.g. Forman & Jones 1982; Gioia et al. 1990), followed by studies with the ROSAT all-sky survey (e.g. Ebeling et al. 2000; Borgani et al. 2001; Böhringer et al. 2004, 2017; Henry et al. 2009; Klein et al. 2018). By simply counting the number of observed galaxy clusters one can confront cosmological model predictions and survey observations. However, it has been established that observational selection effects play a crucial role and must be controlled accurately when pursuing the goal of precision cosmology (e.g. Vikhlinin et al. 2009; Mantz et al. 2010b; Allen et al. 2011; Pacaud et al. 2016).

X-ray astronomy will enter a new era with the extended ROentgen Survey with an Imaging Telescope Array (eROSITA, Predehl 2017). This telescope is the primary instrument of the Russian/German Spektrum-Roentgen-Gamma (SRG) observatory, expected to be launched in 2019 (P. Predehl, priv. comm.). eROSITA will possess unprecedented sensitivity and imaging capabilities for extended source emission (Merloni et al. 2012), and allow the detection of approximately  $10^5$  galaxy clusters (Pillepich et al. 2012). In order to detect this huge number of galaxy clusters, eROSITA will scan the entire sky for four years,

making it the second imaging X-ray all-sky survey ever made after ROSAT in the soft band (0.5–2 keV), and the first ever imaging survey in the hard band (2–8 keV). The promising capabilities of eROSITA bring great expectations to constrain dark matter and dark energy models through galaxy cluster science.

The derivation of a selection function for extended X-ray sources involves first their detection and then their classification as extended objects. Because extended objects are defined, in contrast to point-like sources, this paper also focuses on the simulation and selection of point-like sources in the eROSITA All-Sky Survey (eRASS).

A reliable detection probability function of point-sources is crucial for assessing the completeness of samples, understanding the X-ray background, evaluating clustering studies, and so on. Given the simple morphology of point-sources, detection probabilities may rely on knowledge of the local exposure time and background levels in a given observation (e.g. Georgakakis et al. 2008). An alternative and common approach consists in simulating mock observations accounting for a range of instrumental and astrophysical effects. Although this method is more computationally demanding, it embraces the entire chain from light emission to source detection and cataloguing, and this is the approach adopted in this work.

As mentioned previously, a selection function for extended sources is a critical ingredient in almost all studies of the X-ray galaxy cluster population, including cosmological studies, scaling relation works (Stanek et al. 2006; Pacaud et al. 2007; Mantz et al. 2010a; Giodini et al. 2013; Lovisari et al. 2015; Andreon et al. 2016), and detailed studies of the evolution of the intra-cluster medium physics and chemistry (see Böhringer & Werner 2010, for a review). The morphological complexity and diversity of the X-ray cluster population makes it more difficult to accurately describe selection effects. Comparison between samples detected at different wavelengths (e.g. Wen et al. 2012; Rozo et al. 2014; Sadibekova et al. 2014; Nurgaliev et al. 2017) allows an understanding of potential selection biases, but does not a priori provide a truth table for source detection. Therefore, Monte-Carlo simulations play an essential role in understanding the entire process leading to a validated galaxy cluster catalog. Reducing the diversity of cluster shapes to a sensible and reduced set of parameters sets limits on the computational demand, and, importantly, allows for a link between theoretical (e.g. mass, redshift, etc.) and observational quantities. Cluster fluxes and apparent sizes are among the most relevant of these observables (Böhringer et al. 2000; Pacaud et al. 2006; Burenin et al. 2007).

Such synthetic simulations are not the unique route to address the selection function of clusters and active galactic nuclei (AGNs). Numerical N-body and hydrodynamic simulations play an increasingly important role in this debate. Indeed, as they become more and more realistic in reproducing the observed sky at multiple wavelengths (e.g. Ragagnin et al. 2016), they offer invaluable support in the understanding of selection biases. However, the still large computational requirements limit their usage for statistical studies.

The aim of this work is to forecast and illustrate realistic selection functions for the eRASS cluster and point-source population. It relies on multiple realisations of selected areas of the eROSITA sky, with X-ray-emitting sources described by controlled parametric inputs. For instance, the galaxy cluster population is uniquely described by its apparent flux and size on the sky. We make a special effort to reproduce the main spectro-photometric features of the extragalactic point-source population (AGNs). For the first time, we process eRASS simulation fields

with the eROSITA source-detection software (preliminary version). We derive realistic detection lists, similar to the real detection lists expected for scientific use. In particular, we explore thresholds needed to distinguish between spurious, point-like, and extended sources and provide, given a chosen set of cuts, a first series of selection functions for point-like and extended sources. We demonstrate their practical usability with a prediction of the distribution of galaxy clusters in the eROSITA sky by means of a forward-modelling approach.

This paper is constructed as follows. We first describe the built-in components of the simulations in Sect. 2, and then we describe the simulation engine at the core of the analysis in Sect. 3. We describe our selected simulation and instrumental setup, as well as our choice of fields in Sect. 4. In Sect. 5 we show the source detection results, in particular the selection functions. We discuss the impact of our important assumptions in Sect. 6 and bring perspectives in Sect. 7.

Throughout the paper and unless stated otherwise, we assume a  $\Lambda$  cold dark matter (CDM) cosmology with  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2. Simulated components

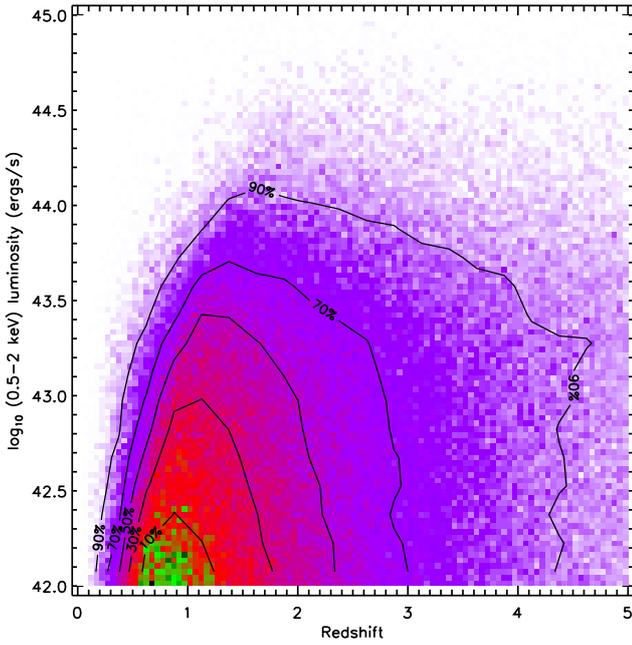
This section presents the main expected components in a typical *blank* field of our simulations of the extragalactic eROSITA sky.

### 2.1. AGN and cosmic X-ray background

We attempt to accurately reproduce the observed distribution of spectro-photometric properties of X-ray-emitting AGNs. A list of spectra and positions, each corresponding to an individual source, is produced down to extremely low fluxes. The integration of the low-flux tail of the distribution provides a model for the unresolved X-ray background component up to the limit at which we simulate sources individually.

#### 2.1.1. Spectral models

We rely on a custom implementation of the formalism by Gilli et al. (2007) to generate spectral models on a log-spaced grid of energies in the range [0.1, 100] keV using XSPEC v12.7.0u (Arnaud 1996). Parameters governing the spectral shape of a source are a power-law photon index,  $\Gamma$ , the absorbing column density,  $N_H$ , the source redshift,  $z$ , and the (unabsorbed) luminosity,  $L_X$ , of the object in a given rest-frame 2–10 keV band. A critical parameter governing the choice of spectral model is the intrinsic absorption  $N_H$ . We call *unobscured* those sources with  $\log_{10} N_H < 21$ , *Compton-thin* those showing  $21 < \log_{10} N_H < 24$ , *Compton-thick mild* those with  $24 < \log_{10} N_H < 25$ , and *Compton-thick heavy* those that have  $\log_{10} N_H > 25$ . For a given obscuration class, two regimes are considered, *Seyfert* or *QSO*, depending on whether the 0.5–2 keV rest-frame luminosity of the source is lower or greater than  $L_X = 10^{46} \text{ erg s}^{-1}$ . We refer to Gilli et al. (2007) for details on the modelling of spectral energy distribution (SED) for each of these classes. The energy range and level of detail in the SED were chosen to match the expected detector performances of eROSITA. Depending on source class, they include a (cut-off) power-law with index  $\Gamma$ , and a 6.4 keV iron line with various equivalent widths (Gilli et al. 1999), possibly modulated by a reflection component. *Compton-thick mild* sources have their cut-off power-law replaced by a more complex p1cabs model (Yaquob et al. 1997). The source is redshifted before applying



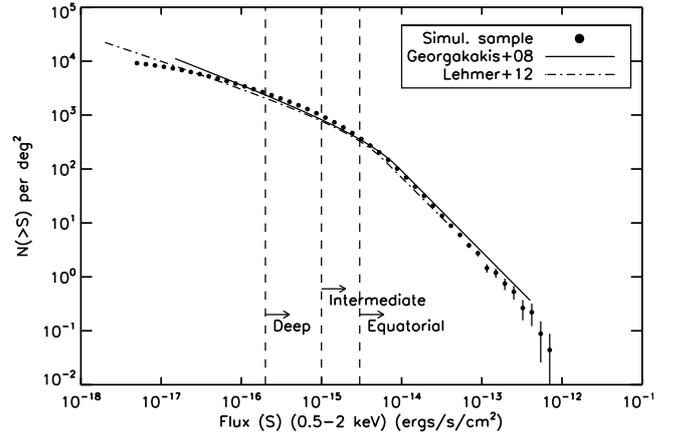
**Fig. 1.** Two-dimensional histogram distribution of simulated sources in one realisation of our X-ray AGN luminosity function sampling for a  $22.7 \text{ deg}^2$  area on the sky (253,297 sources in total). Each black contour encloses the fraction of sources indicated as a label. To each source belongs one X-ray spectral model uniquely defined by the source luminosity, redshift, power-law index  $\Gamma$  and absorbing column density  $N_{\text{H}}$  (Sect. 2.1).

an additional absorption by the Galaxy ( $N_{\text{H}}^{\text{gal}}$ ) depending on the location of the source on the sky. Finally, the flux of a source is obtained by integration of its SED, accounting for the luminosity distance computed in our reference cosmology.

### 2.1.2. Sampling the luminosity functions

Similarly to Gilli et al. (2007) we describe the luminosity function of unobscured AGN sources with the luminosity-dependent density evolution (LDDE) model of Hasinger et al. (2005). Obscured sources are sampled from the LDDE modulated by a multiplicative factor, ranging from four to one as the source intrinsic luminosity increases. Obscuration values are distributed following the prescription by Gilli et al. (2007), while power-law index parameters are drawn from a normal distribution of mean  $\langle \Gamma \rangle = 1.9$  and spread 0.2 regardless of the source obscuration level. Source luminosities range from  $10^{42} \text{ erg s}^{-1}$  and redshifts span the  $0 < z < 5$  interval. After accounting for the cosmological volume, we compute the sky density  $n(\Gamma, N_{\text{H}}, z, L_{\text{X}})$  (units  $\text{deg}^{-2}$ ) and random-sample this distribution in order to obtain a discrete list of sources. Figure 1 represents the density of one such source list in the luminosity-redshift plane. Each source is then assigned an SED as described in the previous section. Sky positions are uniformly distributed in a field, as we do not aim to accurately model the spatial distribution of sources in this work (see Paper II, Ramos-Ceja et al. for a more detailed treatment).

We verified the validity of our sampling procedure by computing the flux distributions of the simulated sources in different bands. We compared our results to Gilli et al. (2007) and to published  $\log N - \log S$ : the agreement in the soft-band is excellent (see Fig. 2), while we predict twice as many heavily obscured sources ( $\log_{10} N_{\text{H}} > 24$ ) in the 2–10 keV band in comparison to Gilli et al. (2007). We attribute this discrepancy for the rarest



**Fig. 2.** Soft-band cumulative source number counts for one realisation of the X-ray AGN luminosity function sampling for a  $22.7 \text{ deg}^2$  area on the sky. Error bars are  $\sqrt{n}$  for each point. The parametrized  $\log N - \log S$  from Georgakakis et al. (2008) and Lehmer et al. (2012) are overplotted (lines) for comparison. Vertical dashed lines indicate the flux  $f_{\text{lim}}$  of the faintest source being simulated in each of the three fields (Sect. 4.3).

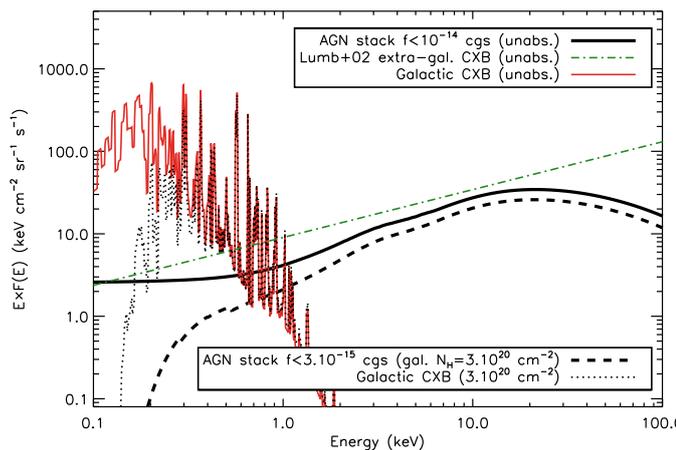
sources to our choices made in the high-energy modeling of the SED. This has practically no impact on this work which focuses on the soft-band characteristics of the eROSITA images.

### 2.1.3. Constructing the unresolved X-ray background

The above-described procedure does not assume a lower limit on the flux of simulated sources. Sources well below the eROSITA detection limit are actually not simulated in order to save computation resources. A flux threshold  $f_{\text{lim}}$  is set depending on the exposure time of a simulated field (Sect. 4.3) and only sources with  $f > f_{\text{lim}}$  are individually simulated. The spectra of the remaining faint sources are stacked together and uniformly redistributed over a simulated patch of sky, thereby constituting one single “uniformly extended source” instead of many point-sources. By doing so, we ensure self-consistent and realistic modelling of the spectral emission of the X-ray background (XRB) generated by unresolved AGNs. As an illustration, the spectrum of the AGN background component in the equatorial field ( $f_{\text{lim}} = 3 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^2$ ) in the 0.5–2 keV band with galactic absorption  $N_{\text{H}}^{\text{gal}} = 3 \times 10^{20} \text{ cm}^{-2}$  is shown with a dashed line in Fig. 3. This figure also demonstrates the good agreement between the *XMM-Newton* measurements of (Lumb et al. 2002; derived from *XMM-Newton* observations with sources excised down to  $\sim 10^{-14} \text{ erg s}^{-1} \text{ cm}^2$  in the soft-band) and our unresolved XRB model with a similar  $f_{\text{lim}}$ .

## 2.2. Extended sources as $\beta$ -models

Galaxy clusters are simulated in the simplest way using spherically symmetric  $\beta$ -models (Cavaliere & Fusco-Femiano 1978) with different fluxes and core radii values and  $\beta = 2/3$ . Our goal is indeed to derive selection functions that depend on a limited number of parameters. Sources representing galaxy clusters are randomly distributed across a simulated field, with a density of around 2 per  $\text{deg}^2$ . Their spectral emission is rendered by an isothermal APEC model with  $0.3 Z_{\odot}$  abundance, at temperature  $T \in \{1, 5\} \text{ keV}$  and redshift  $z \in \{0.3, 0.8\}$ . Clusters have 0.5–2 keV fluxes chosen among discrete values ranging between  $2 \times 10^{-15}$  and  $5 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^2$ ; core radii are also picked among



**Fig. 3.** Energy spectrum of the simulated cosmic X-ray background components. The thick black dashed and plain lines are obtained with our model for AGN sources by stacking individual spectra of faint sources below  $f_{\text{lim}} = 10^{-14}$  (plain) or  $3 \times 10^{-15}$  (dashed)  $\text{erg s}^{-1} \text{cm}^{-2}$ . For comparison, the dot-dashed green line shows the model of [Lumb et al. \(2002\)](#) derived from *XMM-Newton* observations. Our emission model for the Galaxy (red and dotted lines) is described in the text.

discrete values ranging between 10 and 80 arcsec. The redshift and temperature of the spectral models have practically no impact on the 0.5–2 keV energy conversion factors transforming fluxes into count-rates, and therefore have no impact on the 0.5–2 keV detection tests, which are the core of this study.

### 2.3. Particle and galactic background components

In addition to the X-ray background originating from unresolved AGNs in the field, two other main background components were added to our set of simulations. The contribution of unresolved galaxy clusters and groups to the eROSITA soft-band background is neglected, since it is a small component in the energy and sensitivity regimes relevant to this study (e.g. [Gilli et al. 1999, 2007; Kolodzig et al. 2017](#)).

Following [Lumb et al. \(2002\)](#), the emission of the Galaxy is modelled with a double MEKAL model of temperatures 0.21 and 0.081 keV and solar abundance, representing the emission of the hot plasma located in the Galactic disk and halo. We assume a local photo absorbing column density equivalent to that of the field under consideration. We neglect here any spatially dependent contribution to the Galactic background such as emission from the Hot Local Bubble.

Particle background is sampled from a list of events drawn from a GEANT4 simulation designed to reproduce the expected radiation environment at the Lagrange point L2 ([Tenzer et al. 2010](#)). We assume this background component is not focused by the telescope mirror systems, and therefore it is not vignettted and impacts the detectors uniformly. Soft proton flares can create rapid enhancement of the level of unvignettted background. However, we limit our present study to the case of nominal particle background level and defer the analysis of the flare-induced background to further work. Therefore, the exposure assumptions in this work are on the optimistic side.

## 3. The eROSITA simulation engine

The simulations presented in this paper result in realistic eROSITA-calibrated event lists, similar to those expected to be delivered by the eROSITA ground segment. Such event lists

contain the arrival time and CCD coordinates of the incoming events (photons or particles), as well as a reconstruction of their sky location and absolute energy. We reconstruct these characteristics assuming perfect knowledge of the calibration and spacecraft attitude. We make use of the Monte-Carlo simulator SIXTE<sup>1</sup>. This simulator virtually implements a realistic transfer function converting sky photons into detector events, accurately accounting for CCD characteristics (including response functions and clocking) and telescope mirror behaviour. In order to save computation time, some parts of the telescope+instrument transfer function are modelled statistically, thus deviating from a pure ray-tracing simulator. These simplifications show notably at the mirror (point-spread and vignetting functions) and the CCD (response function) stages. We refer to [Schmid \(2012\)](#) for a detailed description of the SIXTE and its implementation in the context of eROSITA.

The detectors were simulated assuming an integration time of 50 ms and a finite readout time of the 384 CCD lines (pile-up effects are not relevant in this work). Response matrices are taken from rescaled EPIC-pn response matrices; those are of sufficient accuracy here, as we are focusing on broad-band properties. The field-of-view of each of the seven detectors is circular with a diameter of 1.02 deg, corresponding to the extent of the  $384 \times 384$  pixel cameras with pixel size  $9.6''$ .

## 4. Instrumental and observational setup

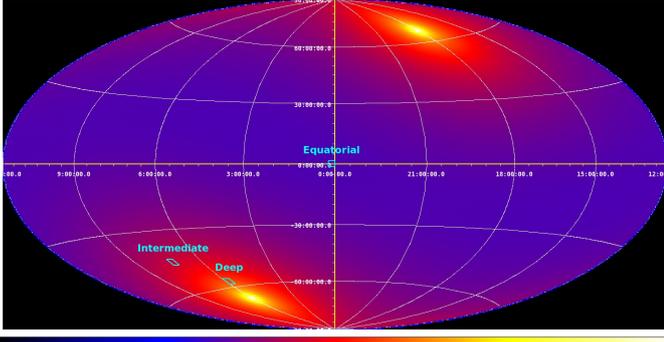
### 4.1. Exposure maps and attitude files

A simple scanning strategy for the four-year survey is assumed in this work, with the spacecraft scanning axis always pointing towards the Sun. The actual spacecraft law will be subject to subtle changes in the scanning pattern in order to fulfill angular constraints linked to, for example, the solar panels or stray-light requirements. Those ultimately lead to less uniform all-sky exposure maps, as discussed in [Merloni et al. \(2012\)](#). Since the present paper focuses on small patches of sky sufficiently far away from the ecliptic poles, these differences are neglected. Extrapolation of our results to the all-sky survey needs, in principle, a proper treatment of these exposure variations. The corresponding attitude files describing the coordinates of the scanning axis in steps of 60 s serve as input to the simulator. We assumed no gaps or jumps over the full duration of the survey, as well as ideal reconstruction of the attitude from the on-board star trackers.

### 4.2. Point-spread function and vignetting

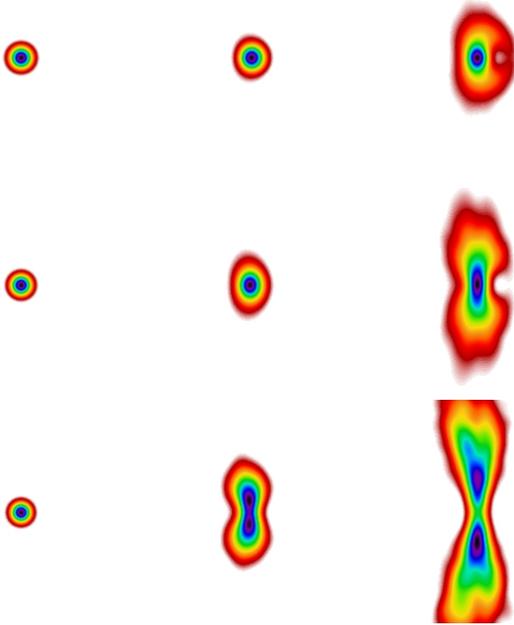
During the simulation procedure, photons originating from a source at infinite distance are redistributed using synthetic point-spread functions simulated with a ray-tracing procedure (P. Friedrich, priv. comm.). This accurately reproduces an eROSITA ideal mirror system made of 54 nested shells (Wolter-I configuration), including the spokes and the presence of an X-ray baffle. Such simulations were performed assuming a focal length of 1.6 m and a 0.4 mm intra-focal shift of the detector relative to the best on-axis focal point. This small shift was found to optimize the overall survey PSF size, at the cost of degrading the on-axis PSF. We note that the actual point-spread function will be measured on the sky when the instrument operates and will be compared to ray-tracing simulations and ground measurements (e.g. as done at the PANTER facility).

<sup>1</sup> <http://www.sternwarte.uni-erlangen.de/research/sixte/>



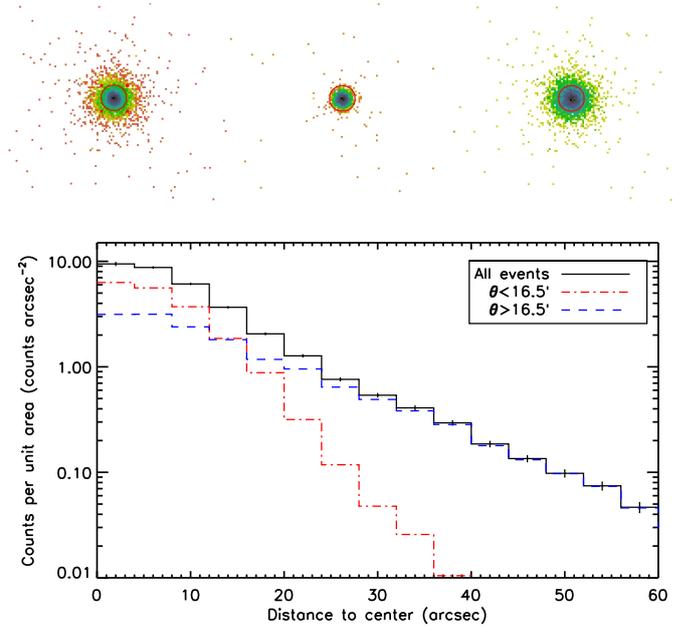
**Fig. 4.** Simulated eROSITA all-sky four-year exposure map in equatorial coordinates used in this work, with the location of the three relevant simulated fields: equatorial ( $\sim 2$  ks exposure time, uniform), intermediate ( $\sim 4$  ks, slight gradient), and deep ( $\sim 10$  ks, larger gradient). The colour bar (logarithmic scale) is in units of seconds.

### 1 arcmin



**Fig. 5.** Ray-tracing simulated telescope point-spread function used in this paper. The images show the response of one eROSITA mirror module to a point-source at different incoming photon energies (from top to bottom: 1, 3 and 7 keV) and different angular distances from the optical axis (from left to right: on-axis,  $15'$ ,  $25'$ ). The colour scale in each panel is linear and encompasses the tenfold increase between the minimal (light red) and maximal (black) intensity, thereby emphasising the typical shape distortions due to Wolter optics.

The PSF we used is described as a tabulated series of images in steps of  $1'$  off-axis angles ranging from 0 to  $30'$  and for energies  $E = \{1, 2, 3, 4, 7\}$  keV (see Fig. 5). We assume constant PSF shape as a function of azimuthal angle, as we consider only axial rotation, as is usual with the Wolter-I telescope symmetry. Because it is counting photons individually, the ray-tracing simulation additionally provides an estimate of the vignetting factor on a grid of energies and off-axis angles. It is used to compute the ratio of flux between double-reflected photons and all photons emitted by a source located at a given off-axis angle, and usually expressed relative to the on-axis position. Figure 6 shows the combined effect of vignetting and PSF distortion on a bright point source passing about 50 times through the eROSITA field-of-view during the four-year scan duration.



**Fig. 6.** Top panel: simulation of a bright point-source with flux  $10^{-11}$  erg s $^{-1}$  cm $^{-2}$  in a four-year eROSITA equatorial region ( $\sim 2$  ks exposure time). The image shows the sky projection of the 0.5–2 keV source events collected by the seven CCD, binned with  $4''$  pixels. Left: “Survey PSF”, including all events. Middle: selecting only low off-axis events ( $\theta < 16.5'$ , 40% of the total number of events). Right: selecting only large off-axis events ( $\theta > 16.5'$ , 60%). The circle has a radius of  $30''$ , slightly larger than the half-energy width of the survey PSF. Bottom panel: corresponding radial profiles in  $4''$  bins (error bars are only shown for the top curve).

**Table 1.** Global parameters for the three types of fields simulated in this study.

	Equatorial	Intermediate	Deep
$T_{\text{exp}}$	$\sim 1.6$ ks	$\sim 4$ ks	$\sim 9.7$ ks
$\Delta T_{\text{exp max}}$	0%	16%	33%
$N_{\text{H}}^{\text{gal}}$ (cm $^{-2}$ )	$3 \times 10^{20}$	$8.8 \times 10^{20}$	$6.3 \times 10^{20}$
$f_{\text{lim}}$ (cgs)	$3 \times 10^{-15}$	$10^{-15}$	$2 \times 10^{-16}$

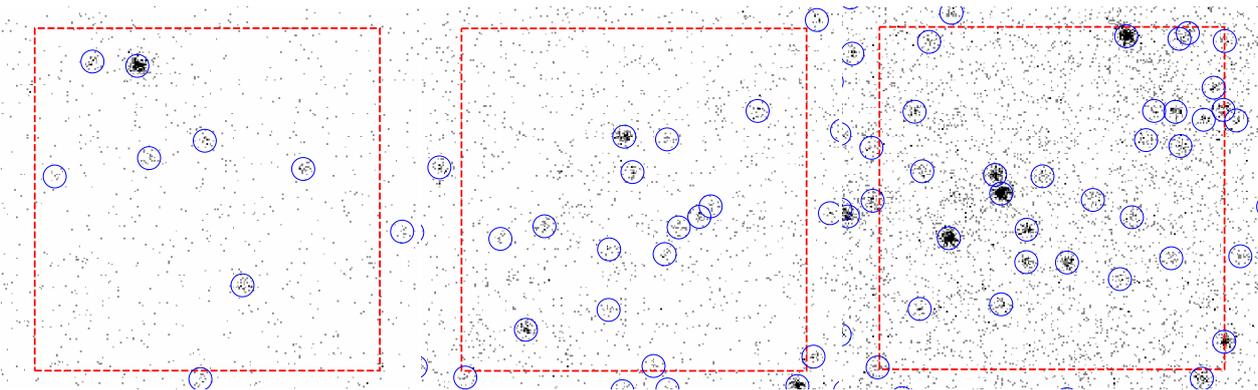
**Notes.** Each field is a square of  $3.6 \text{ deg} \times 3.6 \text{ deg}$ . The galactic absorption is assumed uniform with a value  $N_{\text{H}}^{\text{gal}}$ . AGNs are simulated individually down to a flux  $f_{\text{lim}}$  in the 0.5–2 keV band and sources below  $f_{\text{lim}}$  contribute to a diffuse background component. The maximal variation of exposure across a field is listed as  $\Delta T_{\text{exp}} = (T_{\text{max}} - T_{\text{min}})/T_{\text{mean}}$ .

### 4.3. Simulated fields

We selected three fields at specific locations in the eROSITA sky (see Fig. 4). A field corresponds to an elementary region of the eROSITA sky tiling pattern, and shows as a  $3.6 \text{ deg} \times 3.6 \text{ deg}$  square in tangential projection. In the following we name these fields: Equatorial ( $\sim 2$  ks exposure time, uniform), Intermediate ( $\sim 4$  ks, less uniform), and Deep ( $\sim 10$  ks, larger exposure gradient). Table 1 provides key parameters relevant to these simulated fields.

### 4.4. Images of the synthetic fields

Figure 7 shows a  $15' \times 15'$  excerpt of the images created out of the simulated event lists of *blank fields*, that is, fields without galaxy clusters. The distribution of point-like sources is uniform



**Fig. 7.** Zoom over three simulated eROSITA extragalactic survey fields (equatorial, intermediate and deep, from *left to right*) in the 0.5–2 keV band, free from galaxy clusters (i.e. containing only backgrounds and AGNs as point-sources). North is up and east left; each dashed square has sides of 15' in length. The blue circles have a radius of 0.5' and show the position of the detected sources. The pixel scale is 4'' and identical grey scales are applied to each image to emphasise the differences in sensitivity.

over the sky: any slight apparent gradient in source concentration is an effect of varying exposure times across the fields. The increase in sensitivity clearly makes more sources visible by eye; this figure also outlines the excellent angular resolution of eROSITA, well-adapted to beat confusion effects over most of the survey area, even in deep fields.

## 5. Results

### 5.1. Source detection and characterisation

The source detection and characterisation procedure used in this work is a preliminary version of the source detection tool in the eROSITA Science Analysis Software System (eSASS) package. It builds upon the source detection algorithm used in the *XMM-Newton* Science Analysis System (XMM-SAS) with several revisions and upgrades. The detection procedure is based on the sliding-cell method. As a first step, this algorithm scans an X-ray image with a sliding square box, and if the signal-to-noise ratio (S/N) in the box is greater than a specified threshold value it is marked as a source candidate. The signal is calculated from the pixel values inside the cell, and the background is estimated from the neighbouring pixels. Subsequently, the candidate objects are removed from the image creating a source-free image which is interpolated by a spline function to create a smooth background map. The algorithm convolves the input image with a  $9 \times 9$  pixel ( $36 \times 36$  arcsec) kernel described by a  $\beta = 2/3$ -profile with  $r_c = 15$  arcsec, which roughly matches the survey PSF. The convolved image and the corresponding background map are then used to calculate an S/N map, in which the significant peaks are the positions of the detected sources. In order to increase the sensitivity for large extended sources, this procedure is repeated for  $2 \times 2$  and  $4 \times 4$  rebinned images corresponding to kernels with  $r_c = 30$  and  $r_c = 60$  arcsec, respectively.

Each source candidate identified by the sliding cell algorithm is further analysed by a maximum likelihood fitting method. This technique compares the spatial distribution of the input sources with a PSF<sup>2</sup> convolved with a source extent model ( $\beta$ -profile). The final log-likelihood is calculated by varying the input source parameters, i.e. position, counts, and extent. A multi-PSF fit is also implemented which helps in deblending and

reconstructing the parameters of close-by sources. In the output list, only sources with a log-likelihood above a given threshold are kept.

Among the maximum likelihood fit, output parameters of interest are: *i) detection log-likelihood*, which gives the significance of the detection; *ii) extent*, which is the apparent extension of the best fitting  $\beta$ -model in pixel units; and *iii) extension log-likelihood*, which compares the significance of the extended model and the point-like model. This last parameter classifies the detected sources as point-like (value equal zero) or as extended-like (value greater than zero).

Given that the PSF fitting of the maximum likelihood fitting method is more sensitive to the core of the PSF when on- and off-axis photons are separated, two images from the same simulation and covering the same sky region are produced with photons chosen according to their position on the FoV. The photons are split into inner photons ( $<16.5'$ ) and outer photons ( $>16.5'$ ). In this way, the source detection pipeline runs simultaneously over two images (see Fig. 6).

All simulated images were analysed with the method described above. The detected sources were cross-identified with the simulation inputs using a matching radius of 28 arcsec for point-like sources and 80 arcsec for extended ones.

### 5.2. Source classification

A trade-off between sample completeness and contamination is inevitable when the source selection function in surveys is estimated. Following a methodology introduced in Pacaud et al. (2006), we explore the output parameter space of the maximum likelihood fitting method by means of our simulations in order to set point-like and extended source classification criteria and to estimate their contamination by spurious and misclassified sources. We define spurious detections as those that cannot be identified with any input source within the search radius, and misclassified sources as those point-sources classified by the pipeline as extended sources or vice versa. We define false detections as a single concept that includes spurious and misclassified detections.

#### 5.2.1. Point-source selection functions

AGNs represent the dominant extra-galactic population at X-ray wavelengths. Although the goal of this work is to determine the

<sup>2</sup> This PSF is based on the ray-tracing simulations with 0.4 mm focus offset (see Sect. 4).

galaxy cluster selection function, the estimation of the point-like detection efficiency and its contamination helps to control the systematics in the detection and characterisation of the extended source population.

First, we restrict ourselves to estimating the false detection rate based on the *blank* field simulations, that is, with point-like sources plus background only. We simulate each field 30 times. We find that a simple threshold in the source detection log-likelihood parameter removes most of the false point-like sources while maintaining a good detection efficiency. We choose a threshold value of 10, obtaining  $\sim 0.1$ ,  $\sim 0.2$ , and  $\sim 1.1$  spurious sources per  $\text{deg}^2$  for the equatorial, intermediate, and deep fields, respectively. Such false detection numbers correspond to  $\sim 0.1\%$ ,  $\sim 0.2\%$ , and  $\sim 0.3\%$  of the average detected sources per  $\text{deg}^2$  in their respective fields.

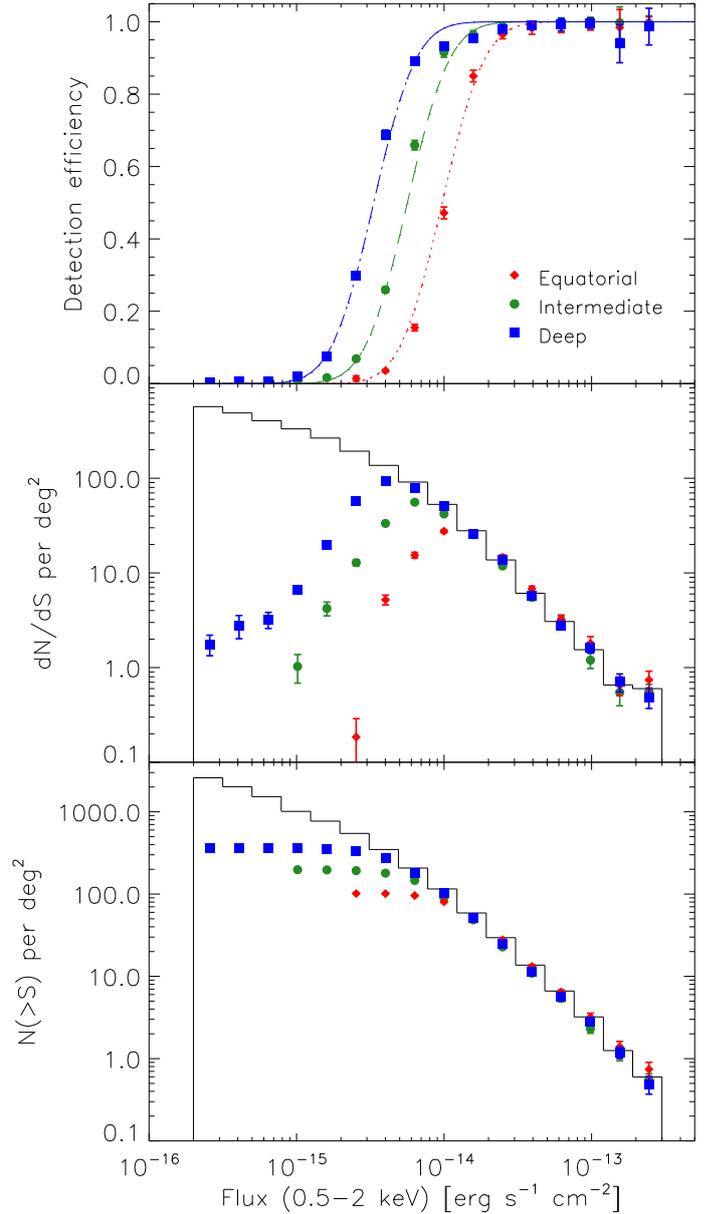
The resulting AGN detection efficiency as a function of input flux is shown in the top panel of Fig. 8. This efficiency is obtained by calculating the ratio of the cross-identified objects to the input sources. The displayed error is given by the standard deviation over the 30 simulations of each simulated field. For the equatorial field, the point-like sources have a 90% completeness at a flux limit of  $\sim 1.7 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ , while for the intermediate field this flux limit is  $\sim 9.7 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ , and for the deep field it is  $\sim 6.5 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ . The 50% completeness is reached at  $\sim 1.0 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$  for the equatorial field,  $\sim 5.2 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$  for the intermediate field, and  $\sim 3.1 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$  for the deep field. The large error bars in bright sources reflect mainly their lower number density, which is given by the AGN  $\log N - \log S$  distribution.

### 5.2.2. Cluster selection functions

The extended source classification is a complicated task since it has not to only deal with spurious detections but also with misclassified point-like sources, that is, point-like sources characterised as extended. Moreover, extended sources usually have a low surface brightness making their detection and characterisation a difficult process. Our goal is to find a location in the detection/characterisation parameter space where the majority of the simulated extended sources are recovered while keeping the contamination level at a reasonable rate. This is of special importance given that the goal of eROSITA is to use galaxy cluster counts to constrain the dark energy. We remind here that in contrast with the AGN population, which was simulated following a  $\log N - \log S$ , sources representing galaxy clusters are randomly distributed across the simulated fields with a density of around 2 per  $\text{deg}^2$  (see Sect. 2).

Besides the source detection log-likelihood values stated in the previous section, we scanned the source extent–extension log-likelihood parameter space to look for criteria that allow us to obtain a large and uncontaminated extended source sample while maintaining a high detection rate. For this, we use *cluster* fields, that is, simulations that contain X-ray background, point-like sources, and extended sources. Figure 9 shows the final selection process in the extent–extension log-likelihood plane for the Equatorial (top), Intermediate (middle), and Deep (bottom) fields.

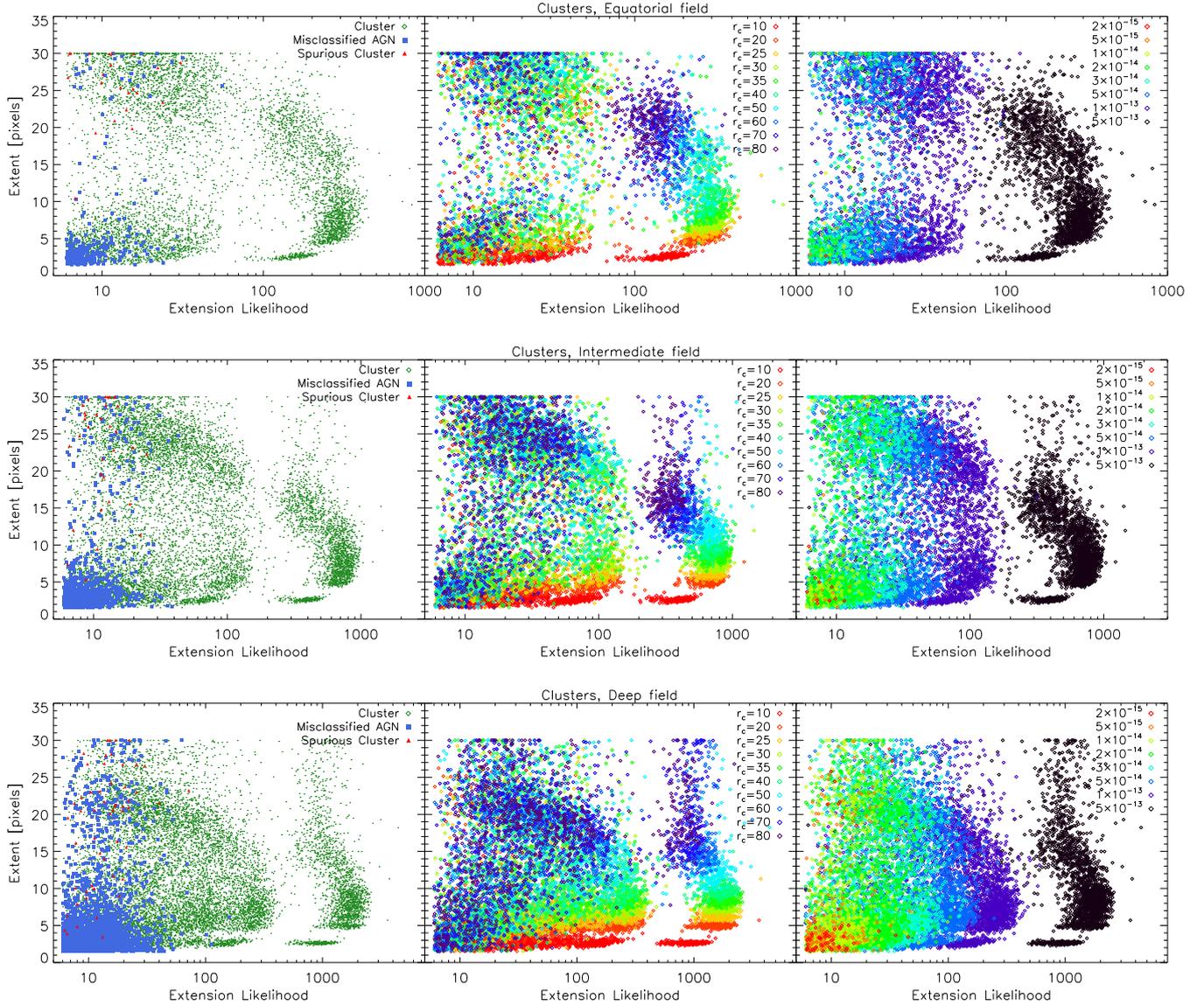
We specify that the maximum extent value that the algorithm should assign to a source is 30 pixels (120 arcsec), even if the algorithm drifts towards a larger value. The minimum requested extent value is 1.5 pixels (6 arcsec), and the threshold of the extension log-likelihood is 6. These thresholds ensure a low contamination by spurious sources, but the number of misclassified point-like sources varies in the different fields. For the equatorial



**Fig. 8.** Point-like source completeness analysis for all three simulated sky regions: Equatorial (red diamonds), Intermediate (green circles) and Deep (blue squares). The abscissa is the input source flux. *Top panel:* point-like detection efficiency. Discontinuous lines represent the parametrized models described in Appendix A. *Middle panel:* differential number counts. *Bottom panel:* integral number of point-like sources. In the *middle* and *bottom panels*, the solid line shows the input distribution. The error is given by the standard deviation over the simulations.

field we obtain  $\sim 0.5$  false extended sources per  $\text{deg}^2$ . In the intermediate field we have  $\sim 1.4$  false extended sources per  $\text{deg}^2$ , and for the deep field we obtain  $\sim 8.5$  false extended sources per  $\text{deg}^2$ .

Table 2 shows in detail the fraction of spurious and misclassified sources in each simulated field. It is worth mentioning that similar numbers of spurious and misclassified sources are found in both the *blank* and *cluster* fields when using the same thresholds. In Sect. 6.2 we forecast the number of expected clusters assuming a survey with a depth equal to the Equatorial field all over the sky. We expect to detect  $\sim 5.2$  clusters per  $\text{deg}^2$  plus 10% contamination from our false sources.



**Fig. 9.** Final selection criteria for extended sources, from the preliminary version of the eSASS pipeline, with optimal (low-contamination) parameters. The extent–extension log-likelihood plane is shown for the three simulated sky fields: equatorial (*top*), intermediate (*middle*), and deep (*bottom*). *Left panels:* simulated (and detected) clusters are displayed as green dots, spurious extended detections as red triangles, and AGNs classified as extended sources in blue squares. *Middle panels:* only input detected galaxy clusters are displayed (green diamonds in the left panels). The distinct colours show the different simulated core radii (in arcsec). *Right panels:* only input detected galaxy clusters are displayed. The different colours show the distinct simulated input fluxes (in units of  $\text{erg s}^{-1} \text{cm}^{-2}$ ).

The middle and right panels of Fig. 9 show the extended sources colour-coded according to the input core radius and flux values, respectively. The middle panels display the distribution of the discrete values used for the core radius of the simulated clusters (see Sect. 2), while the right panels show that mainly sources with high-flux end within the plane of the selection criteria.

As seen in Fig. 9, one could put more stringent criteria to obtain a non-contaminated cluster sample, for example, increasing the minimum value of the extension log-likelihood, but this would lead to excluding a considerable amount of extended sources, especially the faintest ones.

The normalized detection probabilities of extended sources for the three simulated fields are presented in Fig. 10, as a function of the input flux. In these plots, a detection efficiency equal to 1 means that 100% of the simulated sources have been detected and classified as extended. As expected, the deeper

the observation, the fainter the recovered extended sources. Figure 11 shows the mean detection probability of extended sources as a function of input flux and input core radius. Similar to other works (e.g. Vikhlinin et al. 1998; Pacaud et al. 2006; Clerc et al. 2012a), we also found that the extended source detection efficiency is not a function of source flux only, especially for the shallower observations.

## 6. Discussion

### 6.1. Effect of source classification criteria

One could argue that the number of false extended source detections, that is, spurious and misclassified detections, found in the different simulated fields (see Table 2) is high considering that eROSITA will perform an all-sky survey. However, most of the false extended detections are misclassified point-sources.

**Table 2.** Number of spurious and misclassified extended sources (galaxy clusters) and point-like sources (AGN) in the cluster field simulations on the equatorial, intermediate, and deep fields.

	Equatorial field [sources per deg <sup>2</sup> ]	Intermediate field [sources per deg <sup>2</sup> ]	Deep field [sources per deg <sup>2</sup> ]
Spurious extended sources	0.04 ± 0.06	0.05 ± 0.09	0.13 ± 0.09
Spurious point-like sources	0.11 ± 0.08	0.19 ± 0.10	1.00 ± 0.30
Misclassified point-like sources	0.44 ± 0.09	1.38 ± 0.20	8.40 ± 1.80

Such sources might be close pairs of point-sources which cannot be disentangled by the detection algorithm and were therefore classified as an extended source. One way to reduce the number of misclassified sources is by doing a complete follow-up on the detected extended sources. Another way is by putting stricter thresholds in source classification criteria; for example by increasing the extent and extension log-likelihood thresholds (see Sect. 5.2 and Fig. 9). For example, using a threshold value in extension log-likelihood of 20 reduces the number of misclassified point-like sources in the three fields by  $\geq 95\%$ . Although such an approach gives a cleaner sample, many real extended sources are missed.

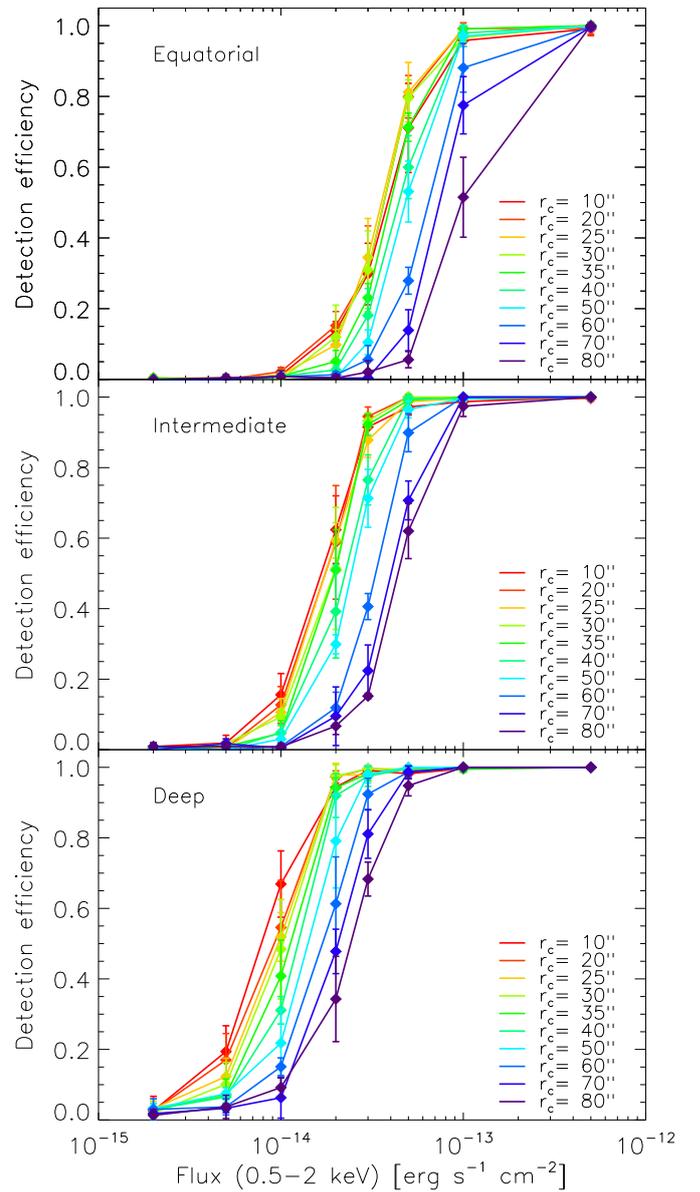
## 6.2. Relevance on cosmological forecasts

Uncertainties in the selection function of a sample of clusters can introduce biases to the cosmological constraints which are determined from them. In this section, we discuss the impact that incomplete knowledge of the selection has on the recovered cosmological constraints. For this test, we follow the methodology of Clerc et al. (2012a) and use the z-CR-HR method. We assume that the selection has eliminated all spurious clusters and misclassified AGNs.

### 6.2.1. The z-CR-HR method

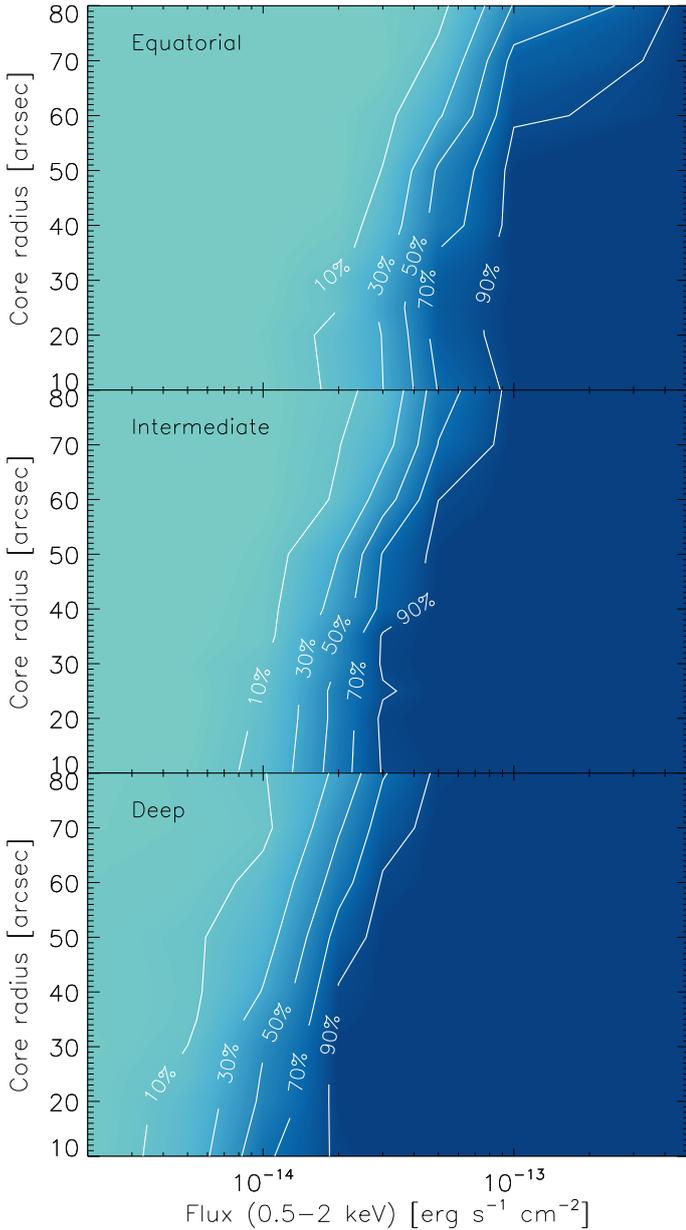
The z-CR-HR method is based on the premise that the raw X-ray data of a galaxy cluster contain significant information about its redshift, luminosity, and temperature and that this information can be statistically extracted. The cosmological analysis is then simplified by basing it on only the cluster redshift and quantities that are directly observable in X-rays, namely the count-rate in the 0.5–2 keV band (CR) and the hardness ratio (HR), which is the ratio of the count-rates measured in the 1–2 keV and 0.5–1 keV bands. A particular advantage of this method is that it bypasses the need to derive individual cluster masses, X-ray luminosities, and temperatures and that the scaling relations between mass and its X-ray proxies can be constrained simultaneously with the cosmological parameters. The key steps in this procedure are as follows:

- Compute the halo mass function.
- Derive the 3D distributions of temperature, luminosity, and core radius using the  $M-T$ ,  $L-T$ , and  $M-r_c$  scaling relations, taking the relevant scatters into account.
- Apply an instrumental model for eROSITA to obtain a theoretical distribution of clusters in the CR-HR plane for each slice in the redshift space.
- Apply the selection function to obtain a synthetic observed distribution of clusters that one would expect eROSITA to detect (here the *equatorial* selection for nominal thresholds, Fig. 10, top).
- Apply an error model to account for measurement errors of CR and HR.


**Fig. 10.** Extended source detection efficiency from the eSASS pipeline in the Equatorial ( $\sim 2$  ks exposure, *top*), Intermediate ( $\sim 4$  ks, *middle*) and Deep ( $\sim 10$  ks, *bottom*) simulated fields as a function of input flux and for each simulated core radius value.

### 6.2.2. Simulated eROSITA z-CR-HR catalogues

After following the procedure described in the previous section and with the unconvolved, error-free z-CR-HR distribution in hand, we randomly sampled the CR-HR plane for each redshift slice to obtain a catalogue of mock clusters each with a redshift,



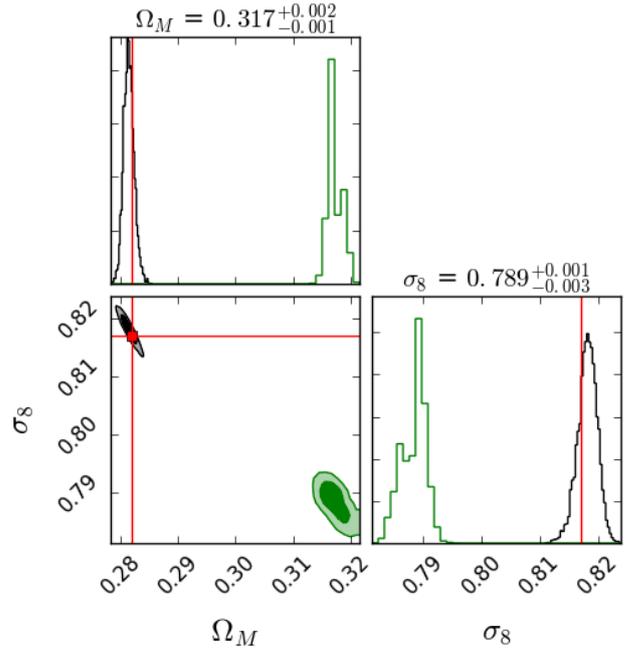
**Fig. 11.** Extended source detection efficiency of the eSASS pipeline in the Equatorial ( $\sim 2$  ks exposure, *top*), Intermediate ( $\sim 4$  ks, *middle*) and Deep ( $\sim 10$  ks, *bottom*) simulated fields as a function of input flux and core radius.

count-rate, and hardness ratio. Gaussian random errors of 10% and 20% for CR and HR, respectively, are then added to each cluster in the mock catalogue<sup>3</sup>. Once the errors have been added, the catalogue is cut with the selection criteria. For this analysis, we apply cuts in CR in  $[0.002, 1]$  cts s<sup>-1</sup> (roughly corresponding to  $[0.28, 140] \times 10^{-14}$  erg s<sup>-1</sup> cm<sup>-2</sup>) and in HR in  $[0.02, 2.0]$ .

### 6.2.3. Cosmological analysis of mocks

In order to recover the input cosmological parameters, we employ a maximum likelihood method and sample the cosmological parameters using a Markov chain Monte Carlo (MCMC)

<sup>3</sup> Although the characterisation of photometric measurements is beyond the scope of this paper, the same simulations as presented in this work can support derivation of such uncertainties.



**Fig. 12.** Bias introduced by the single core radius selection function. The black contours show the recovered constraints from the complete selection function while the green contours are the results obtained by fitting the cosmology assuming a single core radius in the selection function. The contours represent the 68% and 95% confidence intervals, respectively. The red lines indicate the position of the fiducial input values used in the creation of the mock catalogue and the values quoted above the plots indicate the median value recovered when using the incorrect selection function.

method. For the description of the likelihood we make use of the unbinned Cash C-statistic (Cash 1979) which provides a useful way of determining how well a given set of data fits the expected distribution. The log-likelihood which we compute for each set of cosmological parameters is given by,

$$\ln \mathcal{L} = \sum_i \ln \left( \frac{dn}{dCR dHR} (CR_i, HR_i) \right) - \int_{CR_{\min}}^{CR_{\max}} \int_{HR_{\min}}^{HR_{\max}} \frac{dn}{dCR dHR} dCR dHR, \quad (1)$$

where the sum in the above equation runs over all selected clusters and the integral (calculated over the cluster selection criteria) gives the number of clusters expected to be within the CR-HR region.

For this work, we chose to use the publicly available Python package `emcee` (Foreman-Mackey et al. 2013), an affine invariant ensemble sampler.

For this analysis, we assume a  $\Lambda$ CDM cosmological model relying on the parameters calculated by Hinshaw et al. (2013), in particular with  $\Omega_m = 0.28$ ,  $\Omega_\Lambda = 0.72$ ,  $\sigma_8 = 0.82$  and  $H_0 = 70$  km s<sup>-1</sup> Mpc<sup>-1</sup>. The scaling relations for  $M-T$  and  $L-T$  are those derived by the XXL collaboration (Pacaud et al. 2016; Giles et al. 2016; Lieu et al. 2016). We only fit for two cosmological parameters,  $\Omega_M$  and  $\sigma_8$ , since we only wish to show that incomplete knowledge of the selection function results in a bias to the recovered parameters. As shown in Fig. 10, the eROSITA selection function is defined for a series of values for the core radius. Here we consider the effect of assuming a selection function which is defined only for a single value of 35 arcsec for the core radius. This core radius is obtained as a weighted average

of the core radii (in arcminutes) of the X-CLASS sample of clusters, Clerc et al. (2012b) and Ridl et al. (2017).

A total of 104 574 clusters were generated over a hypothetical survey of 20 000 square degrees to a uniform depth of 1.6 ks. The selection criteria for clusters entering the mock were  $0.002 < CR < 1.0 \text{ cts s}^{-1}$  and  $0.02 < HR < 2.0$ . The results obtained from the MCMC likelihood analysis are shown in Fig. 12. We see that very tight and unbiased constraints on both  $\Omega_m$  and  $\sigma_8$  are obtained when the selection function is precisely known, as illustrated by the black contours in Fig. 12. On the other hand, a significant bias (shown by the green contours) is observed for both of these parameters when one assumes a core-radius-independent selection function when attempting to fit the cosmological parameters.

## 7. Conclusions

We have produced and analysed a set of realistic simulations for the eROSITA All-Sky Survey (eRASS) aiming towards precise selection functions for galaxy clusters. Our approach represents a trade-off between realism and tractability, capturing the essential (expected) instrumental and astrophysical features of the eRASS:

- Fields of typical sizes in typical locations of the sky were selected and the exposure maps derived according to the spacecraft scanning law;
- they are populated with AGNs following a realistic spectrophotometric distribution;
- expected X-ray backgrounds (extragalactic and instrumental) are added;
- the instrument is accurately modelled using the SIXTE simulator, combined with accurate ray-tracing PSF and vignetting models as well as a detailed detector model;
- galaxy clusters are simulated with various fluxes and sizes following an average  $\beta$ -model profile.

Our main result consists in a revisited selection function for extended sources defined in the (flux, extent) parameter space. We show that such a selection function can be coupled to cosmological codes and we provide an example with forward-modelling the entire galaxy cluster population with the CR-HR method (Clerc et al. 2012a). Adjusting cosmological parameters to a mock catalog, we demonstrate that inaccurate knowledge of the selection function can lead to a significant bias in the derivation of cosmological parameters.

Such selection functions and results are valid to the extent of our current instrumental and astrophysical knowledge. Refined calibration and measurements (e.g. background, point-spread function, etc.), on-ground and in-orbit, will provide updated results, critically needed for statistical analyses based on the eROSITA all-sky survey. Different source-detection algorithms, possibly combining data from other wavelengths, may result in different quantitative selection functions; however the framework presented in this paper remains valid and can be used to quickly and efficiently assess their ability to provide constraints on cosmological models of structure formation.

*Acknowledgements.* The authors thank the anonymous referee for their suggestions and comments which clearly increased the quality of this paper. MERC acknowledges support by the German Aerospace Agency (DLR) with funds from the Ministry of Economy and Technology (BMWi) through grant 50 OR 1608. THR acknowledges support by the German Research Association (DFG) through

grant RE 1462/6 and the Transregio 33 “The Dark Universe” sub-project B18. The authors thank S. Grandis for useful comments on a preliminary version of this paper.

## References

- Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, 49, 409
- Andreon, S., Serra, A. L., Moretti, A., & Trinchieri, G. 2016, *A&A*, 585, A147
- Arnaud, K. A. 1996, in *Astronomical Data Analysis Software and Systems V*, eds. G. H. Jacoby, & J. Barnes, *ASP Conf. Ser.*, 101, 17
- Böhringer, H., & Werner, N. 2010, *A&ARv*, 18, 127
- Böhringer, H., Voges, W., Huchra, J. P., et al. 2000, *ApJS*, 129, 435
- Böhringer, H., Schuecker, P., Guzzo, L., et al. 2004, *A&A*, 425, 367
- Böhringer, H., Chon, G., Retzlaff, J., et al. 2017, *AJ*, 153, 220
- Borgani, S., Rosati, P., Tozzi, P., et al. 2001, *ApJ*, 561, 13
- Burenin, R. A., Vikhlinin, A., Hornstrup, A., et al. 2007, *ApJS*, 172, 561
- Cash, W. 1979, *ApJ*, 228, 939
- Cavaliere, A. & Fusco-Femiano, R. 1978, *A&A*, 70, 677
- Clerc, N., Pierre, M., Pacaud, F., & Sadibekova, T. 2012a, *MNRAS*, 423, 3545
- Clerc, N., Sadibekova, T., Pierre, M., et al. 2012b, *MNRAS*, 423, 3561
- de Haan, T., Benson, B. A., Bleem, L. E., et al. 2016, *ApJ*, 832, 95
- Gilli, R., Comastri, A., Edge, A. C., Allen, S. W., et al. 2000, *MNRAS*, 318, 333
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Forman, W., & Jones, C. 1982, *ARA&A*, 20, 547
- Georgakakis, A., Nandra, K., Laird, E. S., Aird, J., & Trichas, M. 2008, *MNRAS*, 388, 1205
- Giles, P. A., Maughan, B. J., Pacaud, F., et al. 2016, *A&A*, 592, A3
- Gilli, R., Comastri, A., Brunetti, G., & Setti, G. 1999, *New Astron.*, 4, 45
- Gilli, R., Comastri, A., & Hasinger, G. 2007, *A&A*, 463, 79
- Giodini, S., Lovisari, L., Pointecouteau, E., et al. 2013, *Space Sci. Rev.*, 177, 247
- Gioia, I. M., Henry, J. P., Maccacaro, T., et al. 1990, *ApJ*, 356, L35
- Hasinger, G., Miyaji, T., & Schmidt, M. 2005, *A&A*, 441, 417
- Hasselfield, M., Hilton, M., Marriage, T. A., et al. 2013, *JCAP*, 7, 008
- Henry, J. P., Evrard, A. E., Hoekstra, H., Babul, A., & Mahdavi, A. 2009, *ApJ*, 691, 1307
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, 208, 19
- Klein, M., Mohr, J. J., Desai, S., et al. 2018, *MNRAS*, 474, 3324
- Kolodzig, A., Gilfanov, M., Hütsi, G., & Sunyaev, R. 2017, *MNRAS*, 466, 3035
- Lehmer, B. D., Xue, Y. Q., Brandt, W. N., et al. 2012, *ApJ*, 752, 46
- Lieu, M., Smith, G. P., Giles, P. A., et al. 2016, *A&A*, 592, A4
- Lovisari, L., Reiprich, T. H., & Schellenberger, G. 2015, *A&A*, 573, A118
- Lumb, D. H., Warwick, R. S., Page, M., & De Luca, A. 2002, *A&A*, 389, 93
- Mantz, A., Allen, S. W., Ebeling, H., Rapetti, D., & Drlica-Wagner, A. 2010a, *MNRAS*, 406, 1773
- Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010b, *MNRAS*, 406, 1759
- Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2014, *MNRAS*, 440, 2077
- Merloni, A., Predehl, P., Becker, W., et al. 2012, *ArXiv e-prints* [arXiv:1209.3114]
- Nurgaliev, D., McDonald, M., Benson, B. A., et al. 2017, *ApJ*, 841, 5
- Pacaud, F., Pierre, M., Refregier, A., et al. 2006, *MNRAS*, 372, 578
- Pacaud, F., Pierre, M., Adami, C., et al. 2007, *MNRAS*, 382, 1289
- Pacaud, F., Clerc, N., Giles, P. A., et al. 2016, *A&A*, 592, A2
- Pillepich, A., Porciani, C., & Reiprich, T. H. 2012, *MNRAS*, 422, 44
- Planck Collaboration XXIV. 2016, *A&A*, 594, A24
- Predehl, P. 2017, *Astron. Nachr.*, 338, 159
- Ragagnin, A., Dolag, K., Biffi, V., et al. 2016, *J. Astron. Comput.*, submitted [arXiv:1612.06380]
- Ridl, J., Clerc, N., Sadibekova, T., et al. 2017, *MNRAS*, 468, 662
- Rozo, E., Bartlett, J. G., Evrard, A. E., & Rykoff, E. S. 2014, *MNRAS*, 438, 78
- Sadibekova, T., Pierre, M., Clerc, N., et al. 2014, *A&A*, 571, A87
- Schmid, C. 2012, PhD Thesis, *Uni. Erlangen*
- Stanek, R., Evrard, A. E., Böhringer, H., Schuecker, P., & Nord, B. 2006, *ApJ*, 648, 956
- Tenzen, C., Warth, G., Kendziorra, E., & Santangelo, A. 2010, in *High Energy, Optical, and Infrared Detectors for Astronomy IV, Proc. SPIE*, 7742, 77420Y
- Vikhlinin, A., McNamara, B. R., Forman, W., et al. 1998, *ApJ*, 502, 558
- Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, *ApJ*, 692, 1060
- Wen, Z. L., Han, J. L., & Liu, F. S. 2012, *ApJS*, 199, 34
- Yaqoob, T., McKernan, B., Ptak, A., Nandra, K., & Serlemitsos, P. J. 1997, *ApJ*, 490, L25

## Appendix A: Analytic fit to the point-like and extended source-selection curves

We provide analytic functions that represent the results obtained in Figs. 8 and 10. Due to the limited number of points sampling the curves in the steep transition region, we fitted functions that constitute a reasonable representation of the simulation.

For the extended source selection (galaxy clusters), we parametrize the completeness, dubbed  $c$ , as a function of 0.5–2 keV flux, exposure time ( $T_{\text{exp}}$ ) and core radius ( $r_c$ ) as follows:

$$a(T, R) = 13.5 - (R - 1.2)^2 + (T - 3.204)/1.28$$

$$c(F, T, R) = 0.5 + 0.5 \operatorname{erf}((F + a(T, R))/0.2),$$

where  $\operatorname{erf}$  represents the error function,

$$T = \log_{10}(T_{\text{exp}}/[\text{ks}])$$

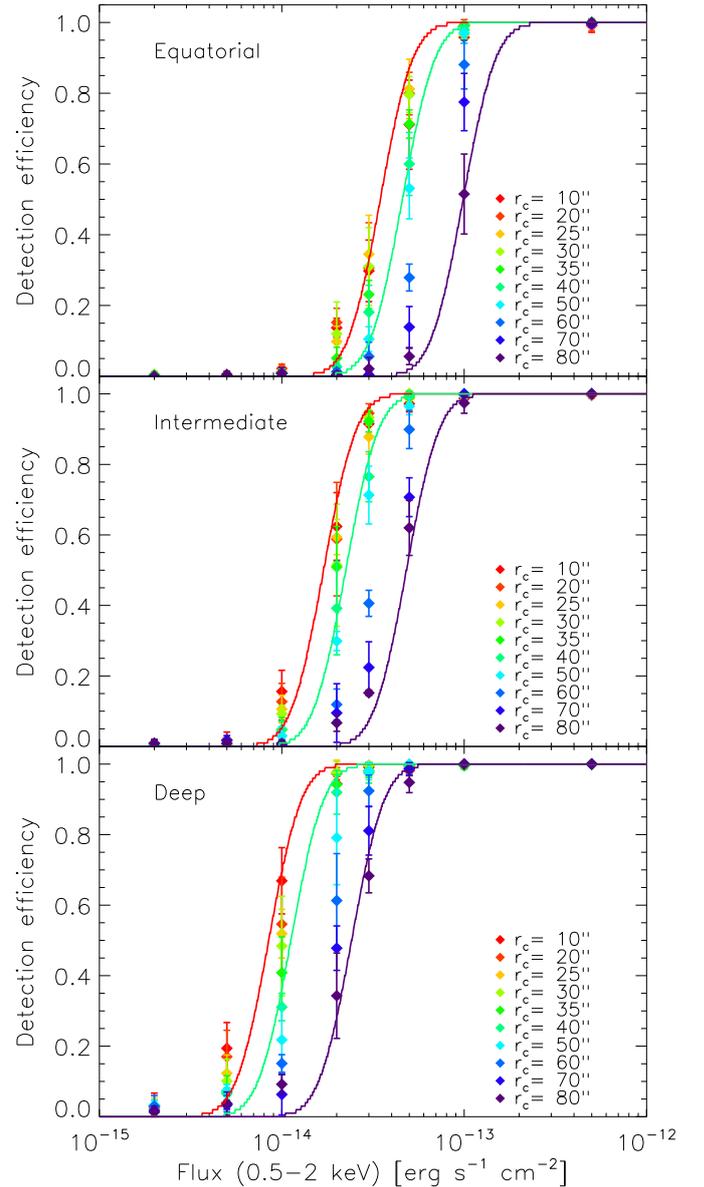
$$R = \log_{10}(r_c/[\text{arcsec}])$$

$$F = \log_{10}(\text{Flux}/[\text{erg cm}^{-2} \text{ s}^{-1}]).$$

For the point-like sources, the parametrization only depends on flux and exposure time:

$$c(F, T) = 0.5 + 0.5 \operatorname{erf}\left(\frac{F + 0.5936(T - 3.1828)}{0.3204}\right). \quad (\text{A.1})$$

In Figs. 8 and A.1 we show the models and their relatively good agreement to the data points extracted from the simulations. Such simple models cannot fully account for the details of the selection function curves, but they should be useful to provide ready-to-use estimates of completeness for various forecasts.



**Fig. A.1.** Similar to Fig. 10, where we superimposed the model lines computed according to formulas in Appendix A, for  $r_c = 10, 40, 80''$ .