Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **498,** 771–798 (2020)
Advance Access publication 2020 August 11

doi:10.1093/mnras/staa2333

Validation of selection function, sample contamination and mass calibration in galaxy cluster samples

S. Grandis,[1,2★] M. Klein,[1,3] J. J. Mohr,[1,2,3] S. Bocquet,[1,2] M. Paulus,[1,3] T. M. C. Abbott,[4] M. Aguena,[5,6] S. Allam,[7] J. Annis,[7] B. A. Benson,[7,8,9] E. Bertin,[10,11] S. Bhargava,[12] D. Brooks,[13] D. L. Burke,[14,15] A. Carnero Rosell,[16] M. Carrasco Kind,[17,18] J. Carretero,[19] R. Capasso,[20] M. Costanzi,[21,22] L. N. da Costa,[23,24] J. De Vicente,[16] S. Desai,[25] J. P. Dietrich,[1] P. Doel,[13] T. F. Eifler,[26,27] A. E. Evrard,[28,29] B. Flaugher,[7] P. Fosalba,[30,31] J. Frieman,[7,32] J. García-Bellido,[33] E. Gaztanaga,[30,31] D. W. Gerdes,[28,29] D. Gruen,[34,14,15] R. A. Gruendl,[17,18] J. Gschwend,[6,24] G. Gutierrez,[7] W. G. Hartley,[13,35] S. R. Hinton,[36] D. L. Hollowood,[37] K. Honscheid,[38,39] D. J. James,[40] T. Jeltema,[37] K. Kuehn,[41,42] N. Kuropatkin,[7] M. Lima,[6,7] M. A. G. Maia,[7,24] J. L. Marshall,[43] P. Melchior,[44] F. Menanteau,[17,18] R. Miquel,[45,19] R. L. C. Ogando,[23,24] A. Palmese,[7,8] F. Paz-Chinchón,[17,18] A. A. Plazas,[44] A. K. Romer,[12] A. Roodman,[14,15] E. Sanchez,[16] A. Saro,[21,22,46] V. Scarpine,[7] M. Schubnell,[29] S. Serrano,[30,31] E. Sheldon,[47] M. Smith,[48] A. A. Stark,[40] E. Suchyta,[49] M. E. C. Swanson,[18] G. Tarle,[29] D. Thomas,[50] D. L. Tucker,[7] T. N. Varga,[3,51] J. Weller[3,51] and R. Wilkinson[12]

Affiliations are listed at the end of the paper

Accepted 2020 August 4. Received 2020 August 4; in original form 2020 March 16

ABSTRACT

We construct and validate the selection function of the MARD-Y3 galaxy cluster sample. This sample was selected through optical follow-up of the 2nd *ROSAT* faint source catalogue with Dark Energy Survey year 3 data. The selection function is modelled by combining an empirically constructed X-ray selection function with an incompleteness model for the optical follow-up. We validate the joint selection function by testing the consistency of the constraints on the X-ray flux–mass and richness–mass scaling relation parameters derived from different sources of mass information: (1) cross-calibration using South Pole Telescope Sunyaev-Zel'dovich (SPT-SZ) clusters, (2) calibration using number counts in X-ray, in optical and in both X-ray and optical while marginalizing over cosmological parameters, and (3) other published analyses. We find that the constraints on the scaling relation from the number counts and SPT-SZ cross-calibration agree, indicating that our modelling of the selection function is adequate. Furthermore, we apply a largely cosmology independent method to validate selection functions via the computation of the probability of finding each cluster in the SPT-SZ sample in the MARD-Y3 sample and vice versa. This test reveals no clear evidence for MARD-Y3 contamination, SPT-SZ incompleteness or outlier fraction. Finally, we discuss the prospects of the techniques presented here to limit systematic selection effects in future cluster cosmological studies.

Key words: methods: statistical – galaxies: clusters: general – large-scale structure of Universe – X-rays: galaxies: clusters.

1 INTRODUCTION

The number of galaxy clusters as a function of mass and redshift is generally accepted as being one of the major sources of information on the composition and evolution of the Universe (see for instance Haiman, Mohr & Holder 2001; Albrecht et al. 2006; Allen, Evrard & Mantz 2011, and references therein). Cluster numbers can be predicted by multiplying the number density of haloes, the 'halo mass function' (HMF), with the effective volume. The HMF is highly sensitive to the matter density and the amplitude of matter fluctuations and can be accurately calibrated by simulations (e.g. Jenkins et al. 2001; Warren et al. 2006; Tinker et al. 2008; Bocquet et al. 2016; McClintock et al. 2019b). The cosmological volume is dependent on the expansion history. Together with the redshift evolution of the amplitude of fluctuations, this makes the redshift evolution of the number of clusters very sensitive to the yet unexplained late time accelerated expansion of the Universe.

Clusters can be selected in large numbers through their observational signatures at different wavelengths. In X-rays, the intracluster medium (ICM), heated by having fallen into the cluster's gravitational potential emits a strong and diffuse thermal emission in X-rays (see Vikhlinin et al. 1998; Böhringer et al. 2001; Romer et al. 2001; Clerc et al. 2014; Klein et al. 2019, for selections based on this signature). At optical wavelengths, clusters can be identified as overdensities of red galaxies (for recent applications

★ E-mail: s.grandis@lmu.de

to wide photometric surveys, see e.g. Koester et al. 2007; Rykoff et al. 2016). In the millimetre regime inverse Compton scattering of the cosmic microwave background (CMB) photons with the ICM makes clusters detectable as extended shadows in CMB maps. This phenomenon is called the Sunyaev–Zel'dovich effect (SZE, Sunyaev & Zeldovich 1972). Scanning CMB surveys for such shadows enables the detection of near-complete, approximately mass-limited cluster samples (Hasselfield et al. 2013; Bleem et al. 2015; Planck Collaboration 2016; Hilton et al. 2018; Huang et al. 2020).

Inference of the mass distribution, and thereby cosmological constraints, is limited by the ability to characterize incompleteness, contamination, and the observable–mass mapping in each of these selection techniques. This leads to the problem of determining the selection function of any cluster sample. The problem is split into two parts: determining how the selection function depends on the X-ray, optical, or SZE observables, and calibrating the scaling relation between that observable and cluster mass. The latter is called *mass calibration* (for a review, see Pratt et al. 2019). It is tackled by measuring the cluster gravitational potential either through the coherent distortion of background galaxies due to weak gravitational lensing (e.g. Bardeau et al. 2007; Okabe et al. 2010; Hoekstra et al. 2012; Applegate et al. 2014; Israel et al. 2014; Melchior et al. 2015; Okabe & Smith 2016; Melchior et al. 2017; Schrabback et al. 2018; Dietrich et al. 2019; McClintock et al. 2019a; Stern et al. 2019), or the analysis of the projected phase space distribution of member galaxies whose velocities are measured by spectroscopic observations (Sifón et al. 2013; Bocquet et al. 2015; Zhang et al. 2017; Capasso et al. 2019a,b). Such techniques are direct probes of the clusters' gravitational potential.

Both X-ray and SZE selections are known to provide cluster candidate lists with less contamination than those carried out in optical surveys which suffer from projection effects (see for instance Costanzi et al. 2019, and reference therein). In X-ray studies at sufficiently high detection significance, extent information can be used to control contamination (Vikhlinin et al. 1998; Pacaud et al. 2006). Nevertheless, optical confirmation is still required to estimate the redshift of the candidates and to reduce the contamination. Traditionally, targeted imaging of individual cluster candidates was performed to this end.

Such campaigns of pointed follow-up have recently been superseded by automated optical confirmation, as for instance by the Multi-wavelength Matched Filter tool (MCMF, Klein et al. 2018). It scans photometric data along the line of sight toward X-ray or SZE cluster candidates with a spatial and colour filter to identify cluster galaxies and determine the clusters redshift. Such tools have the advantage of exploiting the ever larger coverage of deep and wide photometric surveys such as the Dark Energy Survey[1] (DES, Dark Energy Survey Collaboration 2016) or the upcoming *Euclid* Mission[2] (Laureijs et al. 2011) and the Rubin Observatory Legacy Survey of Space and Time[3] (LSST, Ivezic et al. 2019) to follow up candidate lists whose size would make pointed observations impractical.

In this work, we seek to construct and validate the selection function of an X-ray selected and optically cleaned sample. We focus on the MARD-Y3 sample (Klein et al. 2019, hereafter K19) which is constructed by following up the highly contaminated 2nd *ROSAT* faint source catalogue (2RXS, Boller et al. 2016) with the DES data of the first 3 yr of observations (between 2013 August and 2016 February, DES-Y3, DES Collaboration 2018). Strict optical cuts lead to a final contamination of 2.5 per cent at the cost of optical incompleteness, which we model alongside the X-ray selection. The optical follow up also provides measurements of the optical richness of the selected clusters.

First, we confirm the selection function model by constraining the scalings between X-ray flux and mass, and richness and mass in different ways. We perform a cross-calibration of our observables with the indirect mass information contained in the SZE signature of SPT-selected clusters by cross-matching the two samples to derive the flux and richness scaling relation parameters. Then we constrain the same scaling relations by fitting for the number counts of MARD-Y3 clusters while marginalizing over cosmology. The former method is largely independent of the selection function for the MARD-Y3 sample, while the latter is strongly dependent on it. Consequently, consistent scaling relation parameter constraints from the two methods validate our selection function model.

We also test the selection functions by further developing the formalism of cross-matching and detection probabilities introduced by Saro et al. (2015, hereafter S15). We thereby constrain the probability of MARD-Y3 contamination, the SPT-SZ incompleteness and the probability of outliers from the scaling relations. This also allows us to identify a population of clusters that exhibit either a surprisingly high X-ray flux or surprisingly low SZE signal.

The paper is organized as follows. Section 2 presents the conceptual framework within which we model galaxy cluster samples; Section 3 presents the specific validation methods used in this work; and Section 4 contains a description of the MARD-Y3 and the SPT-SZ cluster samples as well as the priors adopted for the analysis. Our main results are presented in Section 5, comprising the different cross-checks on scaling relation parameters that validate our selection function modelling and that enable further checks of the selection functions. The results are then discussed in Section 6, leading to our conclusions in Section 7. The appendices contain more extensive descriptions of the construction and validation of the X-ray observational error model (Appendix A) and a gallery of multiwavelength images used for visual inspection (Appendix B).

2 CONCEPTUAL FRAMEWORK FOR CLUSTER COSMOLOGY ANALYSES

In the following section, we will present in mathematical detail the model of the cluster population used in this work to describe the properties of the cluster samples. This discussion follows the Bayesian hierarchical framework established by Bocquet et al. (2015). The cluster population is modelled by a forward modelling approach that transforms the differential number of clusters as a function of halo mass M_{500c}[4] and redshift z to the space of observed cluster properties, such as the measured X-ray flux $\hat{f}_{\rm X}$, the measured richness $\hat{\lambda}$, and the measured SZE signal-to-noise ξ. This transformation is performed in two steps. First, scaling relations having intrinsic scatter are utilized to estimate the cluster numbers as a function of intrinsic flux, richness, and SZE signal-to-noise ratio. These relations have several free parameters such as amplitude, mass and redshift trends, intrinsic scatter around the mean relation, and correlation coefficients among the intrinsic scatter on different observables. Constraining these free parameters is the aim of this work, as these constraints characterize

[1] https://www.darkenergysurvey.org
[2] http://sci.esa.int/euclid/42266-summary/
[3] https://www.lsst.org/

[4] $M_{500,c}$ is the mass enclosed in a spherical over density with average density 500 times the critical density of the Universe. For sake of brevity we will refer to $M_{500,\,c}$ as M for the rest of this work.

the systematic uncertainty in the observable–mass relations. Second, we apply models of the measurement uncertainty to construct the cluster number density as a function of their measured properties. We also present the modelling of the selection function and of the likelihood used to infer the parameters governing the scaling relations.

2.1 Modelling the cluster population

The starting point of our modelling of the cluster population is their differential number as function of halo mass M_{500c} and redshift z, given by

$$\left.\frac{\mathrm{d}N}{\mathrm{d}M}\right|_{M,z} = \left.\frac{\mathrm{d}n}{\mathrm{d}M}\right|_{M,z} \frac{\mathrm{d}^2V}{\mathrm{d}z\mathrm{d}\Omega}\mathrm{d}z\mathrm{d}\Omega, \tag{1}$$

where $\left.\frac{\mathrm{d}n}{\mathrm{d}M}\right|_{M,z}$ is the HMF describing the differential number density of haloes at mass M and redshift z, as presented by Tinker et al. (2008); $\frac{\mathrm{d}^2V}{\mathrm{d}z\mathrm{d}\Omega}\mathrm{d}z\mathrm{d}\Omega$ is the cosmological volume subtended by the redshift bin $\mathrm{d}z$, and the survey angular footprint $\mathrm{d}\Omega$.

The mapping from halo mass to intrinsic cluster properties is modelled by scaling relations, which are characterized by a mean relation with free parameters and a correlated scatter. The mean intrinsic relations we use read as

$$\langle \ln f_\mathrm{X}\rangle = \ln \frac{L_0\, A_\mathrm{X}}{4\pi d_L^2(z)} + B_\mathrm{X} \ln\left(\frac{M\,h}{M_{0,\mathrm{X}}}\right) + 2\ln\left(\frac{E(z)}{E(z_{0,\mathrm{X}})}\right) + C_\mathrm{X} \ln\left(\frac{1+z}{1+z_{0,\mathrm{X}}}\right) \tag{2}$$

for the rest frame [0.5, 2] keV X-ray flux (hereafter 'X-ray flux', for details see Section 4.1.1 below),[5]

$$\langle \ln \lambda\rangle = \ln A_\lambda + B_\lambda \ln\left(\frac{M\,h}{M_{0,\lambda}}\right) + C_\lambda \ln\left(\frac{E(z)}{E(z_{0,\lambda})}\right) \tag{3}$$

for the richness, and

$$\langle \ln \zeta\rangle = \ln A_\mathrm{SZ} + \ln B_\mathrm{SZ}\left(\frac{M\,h}{M_{0,\mathrm{SZ}}}\right) + C_\mathrm{SZ} \ln\left(\frac{E(z)}{E(z_{0,\mathrm{SZ}})}\right) \tag{4}$$

for the SZE signal-to-noise ratio in a reference field. h is the present-day expansion rate in units of 100 km s^{-1} Mpc^{-1}, and $E(z)$ the ratio between the expansion rate at redshift z and the current day expansion rate. The form of the redshift evolution adopted in equations (3) and (4) has explicit cosmological dependence in the redshift evolution that is not well motivated (see discussion in Bulbul et al. 2019, hereafter Bu19), but we nevertheless adopt these forms for consistency with previous studies (e.g. S15). The pivot points in mass $M_{0,\mathrm{X}} = 6.35\times10^{14} h\,\mathrm{M}_\odot h^{-1}$, $M_{0,\lambda} = 3.\times10^{14}\,\mathrm{M}_\odot h^{-1} = M_{0,\mathrm{SZ}}$, in luminosity $L_0 = 10^{44}$ erg s^{-1}, and in redshift $z_{0,\mathrm{X}} = 0.45$, $z_{0,\lambda} = 0.6 = z_{0,\mathrm{SZ}}$ are constants in our analysis. In contrast, the parameters $A_\aleph$, $B_\aleph$, and $C_\aleph$ for $\aleph \in$ (X, λ, SZ) are free parameters of the likelihoods described in Section 3. These parameters encode the systematic uncertainty in the mass derived from each observable.

The inherent stochasticity in the cluster populations is modelled by assuming that the intrinsic observable scatters lognormally around the mean intrinsic relation. Consequently, given mass M and redshift z, the probability for the intrinsic cluster observables (f_X, λ, ζ) is given by

$$P(f_\mathrm{X},\lambda,\zeta|M,z) = \frac{1}{\sqrt{(2\pi)^3\det\mathbf{C}}}\frac{1}{f_\mathrm{X}\lambda\zeta}\exp\left\{-\frac{1}{2}\Delta\mathbf{x}^T\mathbf{C}^{-1}\Delta\mathbf{x}\right\}, \tag{5}$$

[5]The flux in this form makes explicit the cosmological dependencies due to distances and to self-similar evolution while allowing for departures from that self-similar evolution.

with

$$\Delta\mathbf{x}^T = (\ln f_\mathrm{X} - \langle\ln f_\mathrm{X}\rangle, \ln\lambda - \langle\ln\lambda\rangle, \ln\zeta - \langle\ln\zeta\rangle) \tag{6}$$

and

$$\mathbf{C} = \begin{bmatrix} \sigma_\mathrm{X}^2 & \sigma_\mathrm{X}\sigma_\lambda\rho_{\mathrm{X},\lambda} & \sigma_\mathrm{X}\sigma_\mathrm{SZ}\rho_{\mathrm{X,SZ}} \\ \sigma_\mathrm{X}\sigma_\lambda\rho_{\mathrm{X},\lambda} & \sigma_\lambda^2 & \sigma_\lambda\sigma_\mathrm{SZ}\rho_{\lambda,\mathrm{SZ}} \\ \sigma_\mathrm{X}\sigma_\mathrm{SZ}\rho_{\mathrm{X,SZ}} & \sigma_\lambda\sigma_\mathrm{SZ}\rho_{\lambda,\mathrm{SZ}} & \sigma_\mathrm{SZ}^2 \end{bmatrix}, \tag{7}$$

where $\sigma_\aleph$ for $\aleph \in$ (X, λ, SZ) encodes the magnitude of the intrinsic lognormal scatter in the respective observable, while the correlation coefficients $\rho_{\aleph,\aleph'}$ encode the degree of correlation between the intrinsic scatters on the respective observables. The scatter parameters and the correlation coefficients are free parameters of our analysis.

The assumption of lognormality is motivated theoretically by two facts: the intrinsic observables are strictly larger than zero, and a lognormal scatter is the simplest model. Operationally, it has the added benefit of allowing one to introduce correlated scatter in a well-defined way. Observationally, deviations from lognormality have not been detected (e.g. Mantz et al. 2016 did not measure any significant skewness in several different observable–mass relations). In this work, we introduce a new framework to test lognormality (cf. Section 3.3).

The differential number of objects as a function of intrinsic observables can be computed by applying the stochastic mapping between mass and intrinsic observables to the differential number of clusters as a function of mass, that is

$$\left.\frac{\mathrm{d}^3N}{\mathrm{d}f_\mathrm{X}\mathrm{d}\lambda\mathrm{d}\zeta}\right|_{f_\mathrm{X},\lambda,\zeta,z} = \int \mathrm{d}M\, P(f_\mathrm{X},\lambda,\zeta|M,z)\left.\frac{\mathrm{d}N}{\mathrm{d}M}\right|_{M,z}. \tag{8}$$

In some parts of our subsequent analysis, we do not require the distribution in SZE signal-to-noise ratio. The differential number of objects as a function of intrinsic X-ray flux and richness can be obtained either by marginalizing equation (8) over the intrinsic SZE signal ζ, or by defining $P(f_\mathrm{X},\lambda|M,z)$ just for the X-ray and optical observable by omitting the SZE part,

$$\left.\frac{\mathrm{d}^2N}{\mathrm{d}f_\mathrm{X}\mathrm{d}\lambda}\right|_{f_\mathrm{X},\lambda,z} = \int \mathrm{d}\zeta \left.\frac{\mathrm{d}^3N}{\mathrm{d}f_\mathrm{X}\mathrm{d}\lambda\mathrm{d}\zeta}\right|_{f_\mathrm{X},\lambda,\zeta,z} = \int \mathrm{d}M\, P(f_\mathrm{X},\lambda|M,z)\left.\frac{\mathrm{d}N}{\mathrm{d}M}\right|_{M,z}. \tag{9}$$

2.2 Modelling measurement uncertainties

The intrinsic cluster observables are not directly accessible as only their measured values are known. We thus need to characterize the mapping between intrinsic and measured observables.

For the X-ray flux, we assume that the fractional error on the flux $\hat\sigma_\mathrm{X}$ is the same as the fractional error in the count rate. For each object (i) in our catalogue we can determine

$$P\left(\hat f_\mathrm{X}^{(i)}|f_\mathrm{X}\right) = \frac{1}{\sqrt{2\pi\left(\hat\sigma_\mathrm{X}^{(i)}\right)^2}}\frac{1}{\hat f_\mathrm{X}^{(i)}}\exp\left\{-\frac{1}{2}\frac{\left(\ln\hat f_\mathrm{X}^{(i)} - \ln f_\mathrm{X}\right)^2}{\left(\hat\sigma_\mathrm{X}^{(i)}\right)^2}\right\}, \tag{10}$$

where $\hat f_\mathrm{X}^{(i)}$ is the measured flux, and f_X is the intrinsic flux. For an application described below, it is necessary to know the measurement uncertainty on the X-ray flux for arbitrary $\hat f_\mathrm{X}$, also those fluxes for which there is no corresponding entry in the catalogue. As described in more detail in Appendix A, we extrapolate the relative measurement uncertainty rescaled to the median exposure time in our footprint, creating a function $\hat\sigma_\mathrm{X}^2(\hat f_\mathrm{X}, z, t_\mathrm{exp})$, which in turn allows us

to compute

$$P(\hat{f}_{\rm X}|f_{\rm X},z,t_{\rm exp}) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{\rm X}(\hat{f}_{\rm X},z,t_{\rm exp})}}\frac{1}{\hat{f}_{\rm X}} \times \exp\left\{-\frac{1}{2}\frac{(\ln\hat{f}_{\rm X}-\ln f_{\rm X})^2}{\hat{\sigma}^2_{\rm X}(\hat{f}_{\rm X},z,t_{\rm exp})}\right\}. \quad (11)$$

Following S15, the measurement uncertainty on the optical richness is modelled as Poisson noise in the Gaussian limit, that is

$$P(\hat{\lambda}|\lambda) = \frac{1}{\sqrt{2\pi\sigma^2_\lambda}}\exp\left\{-\frac{1}{2}\frac{(\hat{\lambda}-\lambda)^2}{\sigma^2_\lambda}\right\}, \quad (12)$$

where $\hat{\lambda}$ is the measured richness as opposed to the intrinsic richness λ and $\sigma^2_\lambda = \lambda$. For each intrinsic richness, this distribution in internally normalized. The measurement uncertainty on the SZE signal-to-noise ratio follows the prescription of Vanderlinde et al. (2010), who have determined the relation between measured SZE signal-to-noise ξ and the intrinsic signal-to-noise ratio, as a function of the effective field depth $\gamma_{\rm f}$,[6] namely

$$P(\xi|\zeta,\gamma_{\rm f}) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(\xi-\sqrt{\gamma^2_{\rm f}\zeta^2+3})^2\right\}. \quad (13)$$

2.3 Modelling selection functions

The selection functions in optical and SZE observables are easy to model as the mapping between measured and intrinsic observables is known and the selection criterion is a sharp cut in the measured observable. For the optical case, the removal of random superpositions by imposing $f_{\rm cont} < 0.05$ in the optical follow-up leads to a redshift-dependent minimal measured richness $\lambda_{\rm min}(z)$, as discussed in K19. This leads to the optical selection function or probability that a cluster is selected within the DES data $P({\rm DES}|\hat{\lambda},z)$, which is a step function in measured richness

$$P({\rm DES}|\hat{\lambda},z) = \Theta(\hat{\lambda}-\lambda_{\rm min}(z)), \quad (14)$$

where $\Theta(x)$ is the Heavyside step function with value 0 at $x<0$, and 1 at $x\geq 0$. Using the measurement uncertainty on richness (equation 12), we construct the optical selection function in terms of intrinsic richness λ as

$$P({\rm DES}|\lambda,z) = P(\hat{\lambda}>\lambda_{\rm min}(z)|\lambda) = \int_{\lambda_{\rm min}(z)}^{\infty}{\rm d}\hat{\lambda}\,P(\hat{\lambda}|\lambda), \quad (15)$$

where DES can take the values 'true' (i.e. the cluster *is* in the sample) or 'false' (i.e. the cluster *fails* to make it into the sample).

The SPT catalogue we use in this work is selected by a lower limit to the measured signal-to-noise $\xi > 4.5$, which, analogously to the optical case, is a step function in ξ and leads to an SZE selection function or probability that a cluster is selected within the SPT survey $P({\rm SPT}|\zeta,\gamma_{\rm f})$ given by

$$P({\rm SPT}|\zeta,\gamma_{\rm f}) = P(\xi>4.5|\zeta,\gamma_{\rm f}) = \int_{4.5}^{\infty}{\rm d}\xi\,P(\xi|\zeta,\gamma_{\rm f}). \quad (16)$$

[6]In de Haan et al. (2016), these factors are presented as renormalizations of the amplitude of the SZE-signal–mass relation. Our notation here is equivalent, but highlights that they describe a property of the mapping between intrinsic SZE and measured signal, and not between intrinsic signal and mass.

2.3.1 Constraining the X-ray selection function

The selection criterion used to create the 2RXS catalogue is given by the cut $\xi_{\rm X} > 6.5$, where $\xi_{\rm X}$ is the significance of existence of a source, computed by maximizing the likelihood that a given source is not a background fluctuation (Boller et al. 2016). In the space of this observable, the selection function is a simple step function, as shown in the left-hand panel of Fig. 1. In X-ray studies however, the selection function in the space of intrinsic X-ray flux is traditionally determined by image simulations (Vikhlinin et al. 1998; Pacaud et al. 2006; Clerc et al. 2018). In such an analysis, the emission from simulated clusters is used to create simulated X-ray images or event files, which are then analysed with the same source extraction tools that are employed on the actual data. As a function of intrinsic flux, the fraction of recovered clusters is then used to estimate the selection function $P({\rm X\text{-}det}|f_{\rm X},...)$, shown in the right-hand panel of Fig. 1. This process captures, to the degree that the adopted X-ray surface brightness model is consistent with that of the observed population, the impact of morphological variation on the selection.

In this work, we take a novel approach, inspired by the treatment of optical and SZE selection functions outlined above. This approach is based on the concept that the traditional selection function can be described as a combination of two distinct statistical processes: the mapping between measured detection significance $\xi_{\rm X}$ and measured flux $\hat{f}_{\rm X}$ (transition between left-hand and central panels in Fig. 1), and the mapping between measured flux $\hat{f}_{\rm X}$ and intrinsic flux $f_{\rm X}$ (transition between central and right-hand panels in Fig. 1), that is

$$P({\rm X\text{-}det}|f_{\rm X},...) = \int{\rm d}\hat{f}_{\rm X}P({\rm X\text{-}det}|\hat{f}_{\rm X},..)P(\hat{f}_{\rm X}|f_{\rm X},...), \quad (17)$$

where the second part of the integrand is the description of the measurement uncertainty of the X-ray flux. This mapping is needed to perform the number counts and any mass calibration, so it needs to be determined anyway. Its construction is described in Appendix A. In the case of the Roentgen All Sky Survey (RASS), the first term can be easily computed from the mapping between measured flux $\hat{f}_{\rm X}$ and X-ray significance $\xi_{\rm X}$, $P(\xi_{\rm X}|\hat{f}_{\rm X},..)$. Indeed, it is just the cumulative distribution of that mapping for $\xi_{\rm X} > 6.5$.

The mapping between measured flux $\hat{f}_{\rm X}$ and X-ray significance $\xi_{\rm X}$ can be seen in Fig. 2 for the MARD-Y3 clusters, where we plot the detection significance against the measured fluxes. The relation displays significant scatter, which is partially due to the different exposure times (colour-coded). Also clearly visible is the selection at $\xi_{\rm X} > 6.5$ (black line). As an empirical model for this relation, we make the ansatz

$$\langle\ln\xi_{\rm X}\rangle = \ln\xi_0(z)+\alpha_0+\alpha_1\ln\left(\frac{\hat{f}_{\rm X}}{f_0(z)}\right)+\alpha_2\ln\left(\frac{t_{\rm exp}}{400{\rm s}}\right), \quad (18)$$

where $\xi_0(z)$ and $f_0(z)$ are the median significance and measured flux in redshift bins. To reduce measurement noise, we smooth them in redshift. We then assume that the significance of each cluster scatters around the mean significance with a lognormal scatter σ_α. This provides the distribution $P(\xi_{\rm X}|\hat{f}_{\rm X},z,t_{\rm exp}) = \xi_{\rm X}^{-1}N(\ln\xi_{\rm X}|\langle\ln\xi_{\rm X}\rangle,\sigma^2_\alpha)$.

To fit the free parameters of this relation, namely $(\alpha_0,\alpha_1,\alpha_2,\sigma_\alpha)$, we determine the likelihood of each cluster i as

$$L_{\alpha,i} = \frac{P\left(\xi^{(i)}_{\rm X}|\hat{f}^{(i)}_{\rm X},z^{(i)},t^{(i)}_{\rm exp}\right)}{P\left(\xi_{\rm X}>6.5|\hat{f}^{(i)}_{\rm X},z^{(i)},t^{(i)}_{\rm exp}\right)}, \quad (19)$$

where the numerator is given by evaluating $P(\xi_{\rm X}|\hat{f}_{\rm X},z,t_{\rm exp})$ for each cluster, while the denominator ensures proper normalization for the actually observable data, that is clusters with $\xi_{\rm X} > 6.5$. In properly normalizing we account for the Malmquist bias introduced by the

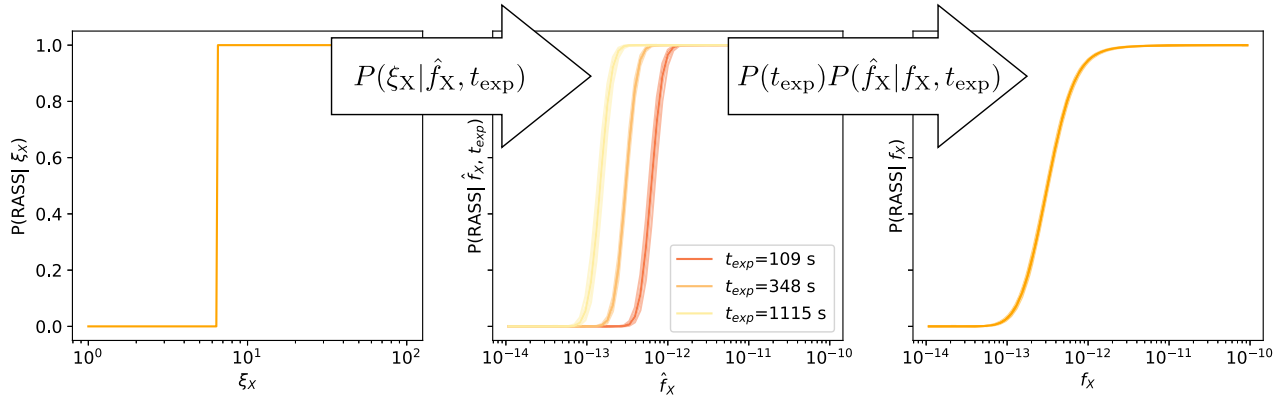

Figure 1. Schematic representation of the construction of the X-ray selection function: in the space of detection significance ξ_X (left), the selection function is simply a step function. Objects above the selection cut are selected, and objects below not. Using the stochastic mapping $P(\xi_X|\hat{f}_X, t_{exp})$ between detection significance and measured flux $\hat{f}_X$ at a given exposure time t_{exp}, we construct the selection function in the space of measured flux $\hat{f}_X$ and exposure time t_{exp} (centre). Accounting for the solid angle weighted exposure time distribution $P(t_{exp})$ and the measurement uncertainty on the flux $P(\hat{f}_X|f_X, t_{exp})$, results in the X-ray selection function in the space of intrinsic flux f_X (right). The latter is usual determined in image simulations making specific assumption on the cluster morphology. In contrast, our method empirically determines the mapping between detection significance and measured flux, thereby accounting for the real morphological variation in the cluster population as well as its uncertainty.

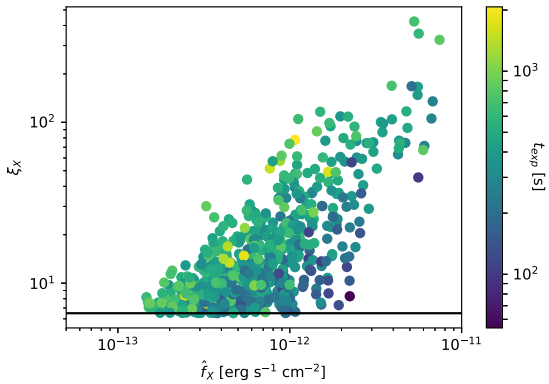

Figure 2. The measured X-ray flux $\hat{f}_X$ and the X-ray detection significance ξ_X, colour coded by the exposure time. The black horizontal line indicates the X-ray selection criterion $\xi_X > 6.5$. While X-ray flux and significance clearly display scaling, the scatter around this scaling correlates with exposure time. This relation and the scatter around it can be used to estimate the X-ray selection function (cf. Section 2.3.1).

X-ray selection. Note also that we do not require the distribution of objects as a function of $\hat{f}_X$ to perform this fit, as it would multiply both the numerator and the denominator and hence cancel out.

The total log-likelihood of the parameters $(\alpha_0, \alpha_1, \alpha_2, \sigma_\alpha)$ is given by the sum of the log-likelihoods $\ln L_\alpha = \sum_i \ln L_{\alpha,i}$. This likelihood provides stringent constraints on the parameters $(\alpha_0, \alpha_1, \alpha_2, \sigma_\alpha)$. We find the best-fitting values $\alpha_0 = -0.113 \pm 0.020, \alpha_1 = 1.275 \pm 0.031$, $\alpha_2 = 0.799 \pm 0.038$, and $\sigma_\alpha = 0.328 \pm 0.012$. Noticeably, the constraints are very tight, indicating that the sample itself provides precise information about this relation.

Given this relation, the X-ray selection function can be computed as

$$P(\mathrm{RASS}|\hat{f}_X, z, t_{exp}) = P(\xi_X > 6.5|\hat{f}_X, t_{exp}, z) = \int_{6.5}^{\infty} \mathrm{d}\xi_X P(\xi_X|\hat{f}_X, t_{exp}, z). \quad (20)$$

Whenever the X-ray selection function is required, we sample the extra nuisance parameters with the ancillary likelihood (equation 19), marginalizing over the systematic uncertainties in this element of the X-ray selection function. Further discussion of the parameter posteriors and their use to test for systematics in the selection function can be found in Section 6.1.

2.3.2 *Testing for additional dependencies*

Empirically calibrating the relation governing the X-ray selection function has three benefits. (1) We take full account of the marginal uncertainty in the X-ray selection function. (2) Compared to image simulation, we do not rely on the realism of the clusters put into the simulation. Indeed, we use the data themselves to infer the relation. Together with the aforementioned marginalization this ensures that we do not artificially bias our selection function. (3) We can empirically explore any further trends of the residuals of the significance–flux relation with respect to other quantities.

Trends of the residuals are shown in Fig. 3, where the residual $\sigma_\alpha^{-1}(\ln \xi_X^{(i)} - \langle \ln \xi_X \rangle^{(i)})$ is plotted against redshift (upper left panel), Galactic hydrogen column density (upper right), background count rate in an aperture of 5 arcmin radius (lower left), and measured extent (lower right). As black dots we show the means of the populations in bins along the x-axis. We find a weak trend with hydrogen column density. For simplicity, we let this trend contribute to the overall scatter σ_α. We find no correlation with the background brightness. There is a clear trend with measured extent, as can be expected for extended sources like clusters. We do not, however, follow up on this trend, as 442 of the 708 cluster that we consider have a measured extent of 0 (due to the large point spread function of RASS).

Disturbingly, we find a trend with redshift which is not captured by our model, as can be seen in the upper left panel of Fig. 3. We tend to underestimate the significance given flux and exposure time at intermediate and high redshifts. This residual systematic manifests itself at different stages in our analysis, and we discuss this as it arises and again in Section 6.1.

3 VALIDATION METHODS

As described above, the selection model for the clusters is specified by the form of the mass–observable relation and the intrinsic and observational scatter of the cluster population around the mean relation.

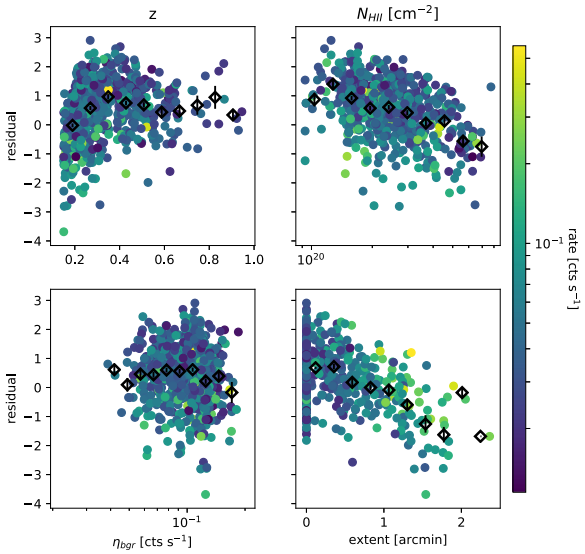

Figure 3. Residuals of the fitted significance–flux relation against redshift (upper left panel), Galactic foreground neutral hydrogen column density (upper right), background counts rate in the aperture (lower left), and measured extent. Colour coded is the count rate of the objects. The residuals are normalized by the best-fitting value of the intrinsic scatter around the mean relation. As black point we show the means of these points in bins along the x-axis. This plot indicates that the addition of a redshift trend or and extent trend would be natural extension of our model. The current level of systematic and statistical uncertainties however does not require these extensions.

The choice of the form of this relation should be driven primarily by what the data themselves demand, with guidance from the principle of preferred simplicity (Occam's razor) and informed by predictions from structure formation simulations. This scaling relation should be empirically calibrated using methods such as weak-lensing and dynamical masses, whose systematics can be calibrated and corrected using by comparison with structure formation simulations. Finally, a key step in cosmological analyses of cluster samples is to check consistency of the cluster sample with the best-fitting model of cosmology and mass–observable relation (e.g. goodness of fit; see Bocquet et al. 2015).

The mass–observable relation can be calibrated using multiple sources of mass information, including direct mass information and from the cluster counts themselves (which is the distribution in observable and redshift of the cluster sample). In future cosmology analyses, blinding of the cosmological and nuissance parameters will be the norm, and cluster cosmology is no exception. The validation of a cluster sample through the requirement that all different reservoirs of information about the scaling relation lead to consistent results can be carried out in a blinded manner and should lead to improved stability and robustness in the final, unblinded cosmological results. We note that given the sensitivity of mass measurements to the distance–redshift relation and the sensitivity of the counts to both distance–redshift and growth of structure, these blinded tests should in general be carried out within each family of cosmological models considered (e.g. flat or curved ν–ΛCDM, flat or curved ν–wCDM, etc).

In this work, we seek to perform the following tests to validate the selection function modelling of the MARD-Y3 sample: (1) we investigate whether the X-ray flux–mass and richness–mass relation obtained by cross-calibration using SPT-SZ mass information is consistent with the relation derived from the number counts of the MARD-Y3 sample; (2) we compare the scaling relation constraints from different flavours of number counts with each other (e.g. number counts in X-ray flux and redshift, in optical richness and redshift, and in both X-ray flux, optical richness, and redshift); and (3) finally, we constrain the probability of incompleteness in the SPT-SZ sample or contamination in the MARD-Y3 sample by comparing the clusters with and without counterparts in the other survey to the probabilities of having or not having counterparts as estimated using the selection functions. We take advantage in these validation tests of the fact that these scaling relations have been previously studied, and so we can compare our results not only internally but also externally to the literature. Finally, a key validation test could be carried out with the weak-lensing information from DES, but we delay that to a future analysis where we hope also to present unblinded cosmological results.

Given the stochastic description of the cluster population outlined above, we set-up different likelihood functions for each of these tests. These likelihoods are functions of the parameters determining the mapping between intrinsic observables and mass, the scatter around these relations and the correlation coefficients among the different components of scatter. Consequently, sampling these likelihoods with the data constrains the parameters. In the following sections, we present the likelihoods used for each of the three validation methods listed above.

3.1 SPT-SZ cross-calibration

For each object in the matched MARD-Y3–SPT-SZ sample (see Section 4.1.3), we seek to predict the likelihood of observing the measured SZE signal-to-noise $\xi^{(i)}$ given the measured X-ray flux $\hat{f}_{\rm X}^{(i)}$, measured richness $\hat{\lambda}^{(i)}$, and the scaling relation parameters. This likelihood is constructed by first making a prediction of the intrinsic SZE-signal-to-noise ζ that is consistent with the measured X-ray flux $\hat{f}_{\rm X}^{(i)}$ and measured richness $\hat{\lambda}^{(i)}$, depending on the scaling relation parameters. To this end, the joint distribution of intrinsic properties is evaluated at the intrinsic fluxes and richnesses consistent with the measurements

$$P\left(\zeta\,|\,\hat{f}_{\rm X}^{(i)},\hat{\lambda}^{(i)},z^{(i)}\right)\propto\int{\rm d}\lambda\,P(\hat{\lambda}^{(i)}|\lambda)\int{\rm d}f_{\rm X}P\left(\hat{f}_{\rm X}^{(i)}|f_{\rm X}\right)\times\frac{{\rm d}^3N}{{\rm d}f_{\rm X}{\rm d}\lambda{\rm d}\zeta}\Big|_{f_{\rm X},\lambda,\zeta,z^{(i)}}.\tag{21}$$

This expression of expected intrinsic SZE signal takes account of the Eddington bias induced by the observational and intrinsic scatter in the X-ray and optical observable acting in combination with the fractionally larger number of objects at low mass, encoded in the last term of the expression.

To evaluate the likelihood of the measured SZE signal-to-noise $\xi^{(i)}$ given the measured X-ray flux $\hat{f}_{\rm X}^{(i)}$ and measured richness $\hat{\lambda}^{(i)}$, we need to compare the predicted distribution $P(\zeta\,|\,\hat{f}_{\rm X}^{(i)},\hat{\lambda}^{(i)},z^{(i)})$ with the likely values of intrinsic SZE signal derived from the measurement $\xi^{(i)}$ and the measurement uncertainty. This is written

$$P\left(\xi^{(i)}|\hat{f}_{\rm X}^{(i)},\hat{\lambda}^{(i)},z^{(i)}\right)=\frac{\int{\rm d}\zeta\;P\left(\xi^{(i)}|\zeta,\gamma_{\rm f}^{(i)}\right)P\left(\zeta\,|\,\hat{f}_{\rm X}^{(i)},\hat{\lambda}^{(i)},z^{(i)}\right)}{\int{\rm d}\zeta\;P\left({\rm SPT}|\zeta,\gamma_{\rm f}^{(i)}\right)P\left(\zeta\,|\,\hat{f}_{\rm X}^{(i)},\hat{\lambda}^{(i)},z^{(i)}\right)},\tag{22}$$

where in the denominator SPT denotes that the cluster *is* detected in the SPT selection. Notably, the denominator ensures the proper

normalization and also takes into account the Malmquist bias[7] introduced by the SPT-SZ selection. Also note that the normalization cancels the dependence of this likelihood on the amplitude of the number of objects at the redshift $z^{(i)}$, measured flux $\hat{f}_{\rm X}^{(i)}$, and measured richness $\hat{\lambda}^{(i)}$. This strongly weakens its cosmological dependence and makes it independent of the X-ray and the optical selection function (see also Liu et al. 2015). For sake of brevity we omitted that this likelihood depends on the scaling relation parameters and the cosmological parameters, all needed to compute the distribution of intrinsic properties.

The total log-likelihood of SPT-SZ cross-calibration over the matched sample is given by the sum of the individual log likelihoods

$$\ln L_{\rm SPTcc} = \sum_{i\in\text{matched}} \ln P\left(\xi^{(i)}|\hat{f}_{\rm X}^{(i)}, \hat{\lambda}^{(i)}, z^{(i)}\right), \tag{23}$$

which is a function of the scaling relation parameters and cosmology. Sampling it with priors on the SZE scaling relation parameters that come from an external calibration will then transfer that calibration to the X-ray flux and richness scaling relations.

3.2 Calibration with number counts

The number of clusters as a function of measured observable and redshift is a powerful way to constrain the mapping between observable and mass, because the number of clusters as a function of mass is known for a given cosmology (see self-calibration discussions in Majumdar & Mohr 2003, 2004; Hu 2003).

3.2.1 X-ray number counts

The likelihood of number counts is given by

$$\ln L_{\rm nc\,X} = \sum_i \ln N\big|_{\hat{f}_{\rm X}^{(i)}, z^{(i)}} - N_{\rm tot}, \tag{24}$$

where the expected number of objects as a function of measured flux $\hat{f}_{\rm X}^{(i)}$ and redshift $z^{(i)}$ is

$$N|_{\hat{f}_{\rm X}^{(i)}, z^{(i)}} = P\left(\text{RASS}|\hat{f}_{\rm X}^{(i)}, z^{(i)}, t_{\rm exp}^{(i)}\right) \int \mathrm{d}f_{\rm X} P\left(\hat{f}_{\rm X}^{(i)}|f_{\rm X}\right) \times \int \mathrm{d}\lambda\, P(\text{DES}|\lambda, z^{(i)}) \frac{\mathrm{d}^2 N}{\mathrm{d}f_{\rm X}\mathrm{d}\lambda}\Big|_{f_{\rm X},\lambda,z^{(i)}} \mathrm{d}\hat{f}_{\rm X}, \tag{25}$$

where the first factor takes into account the X-ray selection, the second factor models the measurement uncertainty on the X-ray flux, and the third factor models the optical incompleteness. As for DES and SPT previously, $P(\text{RASS}|\hat{f}_{\rm X}^{(i)}, z^{(i)}, t_{\rm exp}^{(i)})$ represents the probability that a cluster is selected within the RASS survey.

The total number of objects is computed as

$$N_{\rm tot} = \int \mathrm{d}t_{\rm exp} P(t_{\rm exp}) \int \mathrm{d}z \int \mathrm{d}\hat{f}_{\rm X} P(\text{RASS}|\hat{f}_{\rm X}, z, t_{\rm exp}) \times \int \mathrm{d}f_{\rm X} P(\hat{f}_{\rm X}|f_{\rm X}, z, t_{\rm exp}) \int \mathrm{d}\lambda P(\text{DES}|\lambda, z) \frac{\mathrm{d}^2 N}{\mathrm{d}f_{\rm X}\mathrm{d}\lambda}\Big|_{f_{\rm X},\lambda,z}, \tag{26}$$

where $P(t_{\rm exp})$ is the solid angle weighted exposure time distribution, as determined in K19. We highlight here that, unlike previous work, we explicitly model not only the selection on the X-ray observable, but also fold in the incompleteness correction due to the MCMF optical cleaning via the term $P(\text{DES}|\lambda, z)$.

3.2.2 Optical number counts

While not customary for a predominantly X-ray-selected sample, the number counts of clusters can also be represented as a function of optical richness. In this case, the likelihood reads

$$\ln L_{\rm nc\,\lambda} = \sum_i \ln N\big|_{\hat{\lambda}^{(i)}, z^{(i)}} - N_{\rm tot}, \tag{27}$$

where $N_{\rm tot}$ is given by equation (26), whereas the expected number of clusters as a function of measured richness $\hat{\lambda}^{(i)}$ and redshift $z^{(i)}$ is computed as follows

$$N|_{\hat{\lambda}^{(i)}, z^{(i)}} = \int \mathrm{d}t_{\rm exp} P(t_{\rm exp}) \int \mathrm{d}\hat{f}_{\rm X} P(\text{RASS}|\hat{f}_{\rm X}, z, t_{\rm exp}) \times \int \mathrm{d}f_{\rm X} P(\hat{f}_{\rm X}|f_{\rm X}, z, t_{\rm exp}) \times \int \mathrm{d}\lambda P(\hat{\lambda}^{(i)}|\lambda, z) \frac{\mathrm{d}^2 N}{\mathrm{d}f_{\rm X}\mathrm{d}\lambda}\Big|_{f_{\rm X},\lambda,z} \mathrm{d}\hat{\lambda}, \tag{28}$$

where the first three integrals take account of the X-ray selection, while the last integral models the measurement uncertainty on the richness.

3.2.3 Combined X-ray and optical number counts

Besides determining the number counts in only one observable, one can also determine the number counts in more than one observable (e.g. Mantz et al. 2010), in our case by fitting for the number of objects as a function of both measured flux $\hat{f}_{\rm X}^{(i)}$ and richness $\hat{\lambda}^{(i)}$. We call this flavour of number counts 2D number counts, as opposed to the 1D number counts in either X-ray flux (cf. Section 5.2.1) or richness (cf. Section 3.2.2). The likelihood of 2D number counts reads

$$\ln L_{\rm nc\,X,\lambda} = \sum_i \ln N\big|_{\hat{f}_{\rm X}^{(i)}, \hat{\lambda}^{(i)}, z^{(i)}} - N_{\rm tot}, \tag{29}$$

where the expected number of objects as a function of measured flux $\hat{f}_{\rm X}^{(i)}$ and richness $\hat{\lambda}^{(i)}$ is

$$N|_{\hat{f}_{\rm X}^{(i)}, \hat{\lambda}^{(i)}, z^{(i)}} = P\left(\text{RASS}|\hat{f}_{\rm X}^{(i)}, z^{(i)}, t_{\rm exp}^{(i)}\right) \int \mathrm{d}f_{\rm X} P(\hat{f}_{\rm X}^{(i)}|f_{\rm X}) \times \int \mathrm{d}\lambda P(\hat{\lambda}^{(i)}|\lambda) \frac{\mathrm{d}^2 N}{\mathrm{d}f_{\rm X}\mathrm{d}\lambda}\Big|_{f_{\rm X},\lambda,z^{(i)}} \mathrm{d}\hat{f}_{\rm X}\mathrm{d}\hat{\lambda}, \tag{30}$$

computed by folding the intrinsic number density with the measurement uncertainties on flux and richness.

3.3 Consistency check using two cluster samples

Given the selection functions for two cluster samples, the probability that any member of one sample is present in the other can be calculated. Thus, two distinct tests can be set-up: (1) for each object in the sample A, we can compute the probability of being detected by the sample B, and compare this probability to the actual occurrence of matches; and (2) inversely, we can start from the sample B, compute the probabilities of detection by A, and compare that to the occurrence of matches. This provides a powerful consistency

[7] In cluster population studies redshift information is usually available. Thus, the term 'Malmquist bias' does not refer to the larger survey volume at which high flux objects can be detected, when compared to low flux objects. It refers to that fact that in the presence of scatter among observables, upscattering objects are more likely to pass any selection criterion than downscattered objects. This biases observable–observable plots close to the selection threshold.

check of the two selections functions, and if anomalies are found, this approach can be used, for example, to probe for contamination or unexplained incompleteness in the cluster samples.

3.3.1 MARD-Y3 detection probability for SPT-SZ clusters

For any SPT-SZ cluster with measured SZE signal-to-noise $\xi^{(i)}$ and redshift $z^{(i)}$ in the joint SPT–DES-Y3 footprint we can compute the probability of being detected by MARD-Y3 as follows. We first predict the probability distribution of intrinsic fluxes and richnesses associated with the measured SZE-signal-to-noise ratio as

$$P(f_{\rm X}, \lambda|\xi^{(i)}, z^{(i)}) \propto \int {\rm d}\zeta\, P\left(\xi^{(i)}|\zeta, \gamma_{\rm f}^{(i)}\right) \frac{{\rm d}^3 N}{{\rm d}f_{\rm X}{\rm d}\lambda{\rm d}\zeta}\Big|_{f_{\rm X},\lambda,\zeta,z^{(i)}}. \quad (31)$$

This expression needs to be properly normalized to be a distribution in intrinsic flux and richness. This is achieved by imposing $\int {\rm d}f_{\rm X} \int {\rm d}\lambda\, P(f_{\rm X}, \lambda|\xi^{(i)}, z^{(i)}) = 1$, which sets the proportionality constant of the equation above. Note that this normalization cancels the dependence of this expression on the number of clusters observed.

The predicted distribution of intrinsic fluxes and richnesses needs to be folded with the selection functions to compute the detection probability. The optical selection function is simply given by equation (15) evaluated at the cluster redshift $z^{(i)}$. On the other hand, when computing the X-ray selection function we take the RASS exposure time at the SPT-SZ position into account, while marginalizing over all possible measured fluxes. The X-ray selection function thus reads

$$P\left({\rm RASS}|f_{\rm X}, t_{\rm exp}^{(i)}, z^{(i)}\right) = \int {\rm d}\hat{f}_{\rm X}\, P\left({\rm RASS}|\hat{f}_{\rm X}, t_{\rm exp}^{(i)}, z^{(i)}\right) \times P\left(\hat{f}_{\rm X}|f_{\rm X}, t_{\rm exp}^{(i)}, z^{(i)}\right), \quad (32)$$

where the second factor is taken from equation (11), the expression for the X-ray measurement error at arbitrary measured flux $\hat{f}_{\rm X}$.

The probability of detecting in MARD-Y3 an SPT-SZ cluster with measured SZE-signal-to-noise $\xi^{(i)}$ and redshift $z^{(i)}$ can then be computed by folding the predicted distribution of fluxes and richnesses with the selection functions in flux and richness as follows

$$\begin{aligned} p_{\rm M|S}^{(i)} &:= P({\rm RASS}, {\rm DES}|\xi^{(i)}, z^{(i)}) \\ &= \int {\rm d}f_{\rm X}\, P\left({\rm RASS}|f_{\rm X}, t_{\rm exp}^{(i)}, z^{(i)}\right) \\ &\quad \times \int {\rm d}\lambda\, P({\rm DES}|\lambda, z^{(i)}) P\left(f_{\rm X}, \lambda|\xi^{(i)}, z^{(i)}\right), \end{aligned} \quad (33)$$

where we omit the dependence on the SPT-SZ field depth $\gamma_{\rm f}^{(i)}$ at the position of the SPT-SZ selected cluster.

Given these probabilities, we can define two interesting classes of objects: (1) *unexpected MARD-Y3 confirmations of SPT-SZ detections*, that is SPT-SZ objects that should not have an MARD-Y3 match given their low probability but have been none the less matched, and (2) *missed MARD-Y3 confirmations of SPT-SZ detections*, that is SPT-SZ objects with a very high chance of being matched by MARD-Y3 that have none the less not been matched. For the discussion in this paper we adopt a low probability threshold of $p_{\rm M|S}^{(i)} < 0.025$ for the unexpected confirmations, and we adopt a high probability threshold of $p_{\rm M|S}^{(i)} > 0.975$ for the missed confirmations.

Anticipating that we find a few unexpected MARD-Y3 confirmations and no missed confirmations, we introduce here the probability $\pi_{\rm t}$ that an SPT-SZ cluster that should not be confirmed based on his $p_{\rm M|S}^{(i)}$ is confirmed none the less. The likelihood of $\pi_{\rm t}$ can be computed by following the probability tree shown in Fig. 4. The probability of being matched is $(1 - p_{\rm M|S}^{(i)})\pi_{\rm t} + p_{\rm M|S}^{(i)}$, while the probability of not being matched is $(1 - p_{\rm M|S}^{(i)})(1 - \pi_{\rm t})$. Thus, the log-likelihood is given by

$$\ln L(\pi_{\rm t}) = \sum_{i\in{\rm match}} \ln\left(\left(1 - p_{\rm M|S}^{(i)}\right)\pi_{\rm t} + p_{\rm M|S}^{(i)}\right) + \sum_{i\in{\rm !match}} \ln\left(\left(1 - p_{\rm M|S}^{(i)}\right)(1 - \pi_{\rm t})\right) \quad (34)$$

This likelihood also depends on the scaling relation parameters through the detection probabilities $p_{\rm M|S}^{(i)}$. Marginalizing over these scaling relation parameters accounts for the systematic uncertainty on the observable–mass relations.

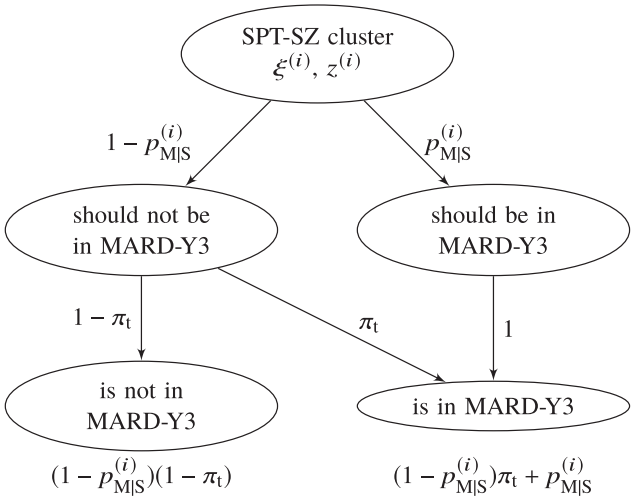

Figure 4. Probability tree describing the probability of an SPT-SZ cluster being detected in MARD-Y3. Besides the matching probability computed from the scaling between intrinsic observables and mass, the scatter around this relation, the observational uncertainties on the observables and the selection functions $p_{\rm M|S}^{(i)}$, we also introduce the chance of either X-ray flux and richness boosting or SZE signal dimming $\pi_{\rm t}$, which would lead to the MARD-Y3 detection of SPT-SZ cluster that should otherwise not have been matched. Summarized at the end of each branch are the probabilities of matching or of not matching.

3.3.2 SPT-SZ detection probability for MARD-Y3 clusters

Similarly to the case in the previous section, for each MARD-Y3 cluster with measured X-ray flux $\hat{f}_{\rm X}^{(i)}$, measured richness $\hat{\lambda}^{(i)}$, and redshift $z^{(i)}$ in the joint SPT-SZ–DES-Y3 footprint, we can compute the probability of it being detected by SPT

$$\begin{aligned} p_{\rm S|M}^{(i)} &:= P\left({\rm SPT}|\hat{f}_{\rm X}^{(i)}, \hat{\lambda}^{(i)}, z^{(i)}\right) \\ &= \int {\rm d}\zeta\, P\left({\rm SPT}|\zeta, \gamma_{\rm f}^{(i)}\right) P\left(\zeta|\hat{f}_{\rm X}^{(i)}, \hat{\lambda}^{(i)}, z^{(i)}\right), \end{aligned} \quad (35)$$

where the first factor in the integral is the SPT-SZ selection function evaluated for the field depth at the MARD-Y3 cluster position, while the second factor is the prediction for the intrinsic SZE-signal-to-noise ratio consistent with the measured X-ray and optical properties. The latter is taken from equation (21) while ensuring that it is properly normalized, $\int {\rm d}\zeta\, P(\zeta|\hat{f}_{\rm X}^{(i)}, \hat{\lambda}^{(i)}) = 1$.

We introduce the probability of each individual MARD-Y3 cluster being a contaminant $\pi_{\rm c}$, and the probability that an SPT-SZ cluster that should be detected has not been detected $\pi_{\rm i}$. From the probability tree shown in Fig. 5 we can determine the probability of a MARD-Y3 cluster being matched by SPT-SZ as $(1 - \pi_{\rm c})p_{\rm S|M}^{(i)}(1 - \pi_{\rm i})$, and

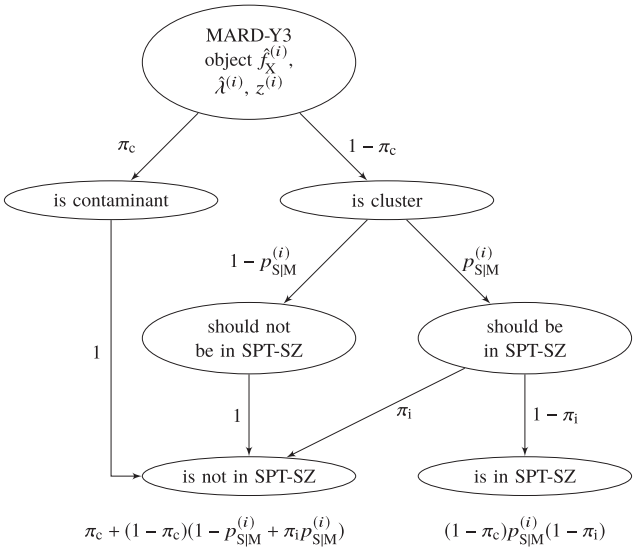

Figure 5. Probability tree describing the probability of an MARD-Y3 cluster being detected by SPT. Besides the matching probability computed from the scaling between intrinsic observables and mass, the scatter around this relation, the observational uncertainties on the observables and the selection functions $p_{S|M}^{(i)}$, we also introduce the chance that an MARD-Y3 cluster is a contaminant π_c, and the chance that SPT-SZ misses a cluster that it should detect, indicating incompleteness in SPT, π_i. Summarized at the end of each branch are the probabilities of being matched or not matched.

the probability of a cluster not being matched as $\pi_c + (1-\pi_c)(1-p_{S|M}^{(i)} + \pi_i p_{S|M}^{(i)})$. Thus, the log-likelihood is given by

$$\ln L(\pi_c, \pi_i) = \sum_{i\in\text{match}} \ln\left((1-\pi_c)p_{S|M}^{(i)}(1-\pi_i)\right) + \sum_{i\in!\text{match}} \ln\left(\pi_c + (1-\pi_c)\left(1-p_{S|M}^{(i)} + \pi_i p_{S|M}^{(i)}\right)\right). \quad (36)$$

This likelihood also depends on the scaling relation parameter through the detection probabilities $p_{M|S}^{(i)}$. Marginalizing over the scaling relation parameters accounts for the the systematic uncertainties on the observable–mass relations. Finally, note that the probability of MARD-Y3 contamination π_c and of SPT incompleteness π_i are perfectly degenerate in this context. We find that the likelihood of SPT confirmation of MARD-Y3 clusters (equation 36) effectively only constrains the difference between the two probabilities. That is $\pi_c = 0.1$ and $\pi_i = 0.0$ is approximately as likely as $\pi_c = 0.0$ and $\pi_i = 0.1$.

3.3.3 Physical interpretation

Several physical effects might bias cluster observables in an unusually significant level compared to the exception from the scatter in observables. In the case of the X-ray flux these effects are, for instance, active galactic nucleus (AGN) contamination and cluster core phenomena. Line-of-sight projections might bias the richness of an object, while extreme astrophysical contamination from correlated radio or dusty emission might bias the SZE signal. The object classes defined above (Sections 3.3.1 and 3.3.2) allow one to select likely candidates for these effects from the comparison of two surveys. This is especially useful in the low signal-to-noise regime where the mass incompleteness of cluster samples in large. In this regime, physical effects within clusters are not resolved. Selecting target lists for high signal-to-noise follow-up might thus further our understanding on the mass incompleteness.

For instance, the classification as an unexpected MARD-Y3 confirmations of an SPT-SZ object can be due to an underestimated detection probability caused by an unexpectedly low SZE signal, or to the X-ray flux and richness being biased high, leading to an actual detection despite the low detection probability. It would thus be indicative of interesting physical properties such as extremely cool cluster cores, strong astrophysical contamination of the SZE signal or strong projection effects in the optical. The presence and impact of these effects would have to be studied with high-resolution X-ray or (sub-)millimetre imaging, or spectroscopic follow-up of the cluster members, respectively. Also note that this class of objects in unlikely to be an MARD-Y3 contaminant, as we find an SPT-SZ object at the same position. Given that the SPT-SZ objects and the putative MARD-Y3 contaminants are both rare on the plane of the sky, the chance of randomly superposing two objects from these classes is small.

As another example, missed MARD-Y3 confirmations of SPT-SZ objects can be due to high SZE signals biasing the detection probability high, or to the X-ray flux and richness being biased low, leading to a non-detection despite the high detection probability. This circumstance is less likely to occur, as astrophysical SZE contaminants usually bias the SZE fluxes low, projection effects bias the richness high, and AGN contamination and cluster core emission bias the X-ray fluxes high. Nevertheless, such an object would be an interesting candidate for an SPT-SZ contaminant, or a case of excess incompleteness in the MARD-Y3 sample. Following the same logic, a missed SPT-SZ confirmation of MARD-Y3 object would indicate either the presence of physical effects that bias the X-ray flux and the richness high, astrophysical contamination that biases the SZE signal low, MARD-Y3 contamination or SPT-SZ excess incompleteness.

4 DATA SET AND PRIORS

We present here the cluster samples and then the priors used in obtaining the results presented in the following section.

4.1 Cluster samples

Here, we summarize not only the main properties of the prime focus of our validation, the MARD-Y3 cluster sample, but also the SPT-SZ sample that we use for validation and for cross-matching with MARD-Y3.

4.1.1 MARD-Y3 X-ray-selected clusters

In this work, we seek to validate the mass information and the selection function modelling of the MARD-Y3 cluster sample, presented in K19. In that work, optical follow-up with the MCMF algorithm (Klein et al. 2018) of the 2RXS (Boller et al. 2016) is performed by scanning the DES photometric data with a spatial filter centred on the X-ray candidate position and inferred mass, and a colour filter based on the red-sequence model at a putative redshift. This process provides a cluster richness estimate $\hat{\lambda}$ and photometric redshift z. Comparison to the richness distribution in lines of sight without X-ray candidates allows one to estimate the probability f_{cont} that the X-ray source and optical system identified by MCMF are a random superposition (contamination). In cases of multiple richness peaks along a line of sight toward a 2RXS candidate, the redshift with lowest f_{cont} is identified as the optical counterpart. The redshifts display subpercent level scatter with respect to spectroscopic redshifts, and the richnesses $\hat{\lambda}$ can be adopted as an additional cluster mass proxy.

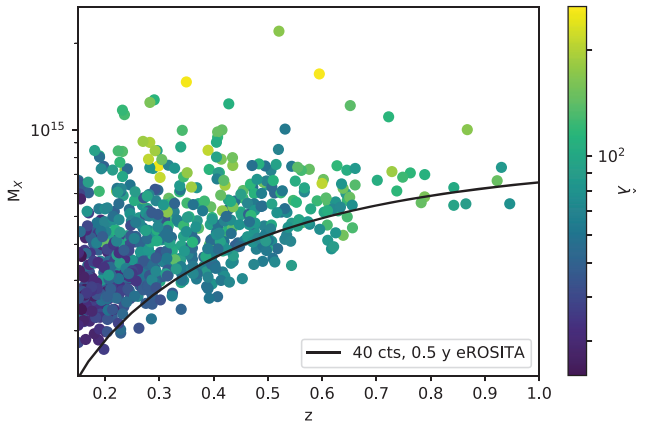

Figure 6. MARD-Y3 sample of 708 clusters constructed by cleaning the 2RXS catalogue with DES data. While not used in the rest of the analysis, the X-ray inferred mass M_X is used here to highlight the mass range of our sample. The colour encodes the measured richness of the counterpart in the DES data. The black line indicates the forecast of the estimated mass corresponding to 40 photon counts in the first eROSITA full sky survey after half a year of observing time.

In this work, we focus on the $z > 0.15$ sample with $f_{\rm cont} < 0.05$ plus an additional rejection of luminosity–richness outliers with an infrared signature compatible with an AGN (cf. K19, section 3.11). Our MARD-Y3 sample is then 708 clusters in a footprint of 5204 deg^2 with an expected contamination of 2.6 per cent (K19).

For these clusters, several other X-ray properties, such as the detection significance ξ_X and the RASS exposure time $t_{\rm exp}$ are available from 2RXS. Boller et al. (2016) originally report luminosities in the [0.1,2.4] keV band extracted from a fixed aperture with radius of 5 arcmin. K19 rescaled these luminosities to luminosities in the rest frame [0.5,2] keV band and within R_{500c}, the radius enclosing an overdensity of 500 wth respect to the critical density. The rescaling is derived from the cross-matching with the meta-catalogue of X-ray detected clusters of galaxies (MCXC) catalogue by Piffaretti et al. (2011). This correction is only reliable at $z > 0.15$. Using the luminosity–mass scaling relation by Bu19 and correcting for Eddington bias, K19 gave an point estimate of the X-ray inferred mass M_X, that we use for plotting purposes. The X-ray flux $\hat{f}_X$ we employ is computed as $\hat{f}_X = L_X/(4\pi d_L^2(z))$, where L_X is the X-ray luminosity within R_{500c}, and $d_L(z)$ is the luminosity distance evaluated at the reference cosmology. This leads to the fact that technically our X-ray flux corresponds to the rest frame [0.5,2] keV. The transformation from the observed [0.1,2.4] keV band to this band is discussed in K19. It is also noteworthy that MCMF allows one to detect the presence of more than one significant optical structure along the line of sight toward an X-ray candidate.

In Fig. 6, we show the redshift–X-ray inferred mass distribution of this sample, colour coded to reflect the cluster richnesses. We also show as a black line the mass corresponding to 40 photon counts in the first eROSITA full sky survey (eRASS1), computed using the eROSITA count rate–mass relation forecast by Grandis et al. (2019). This indicates that the MARD-Y3 sample we study here is comparable to the one we expect to study in the eRASS1 survey.

4.1.2 SPT-SZ SZE-selected clusters

We adopt the catalogue of clusters selected via their SZE signatures in the SPT-SZ 2500 deg^2 survey Bleem et al. (2015). Utilizing this sample to an SZE signal-to-noise ratio of 4.5, we confirm the clusters in the DES-Y3 footprint using MCMF (Klein et al. in preparation). The low contamination level of the parent sample allows one to achieve a low level of contamination by imposing the weak cut of $f_{\rm cont} < 0.2$. Above a redshift of $z > 0.2$ this provides us with a sample of 436 clusters. The X-ray properties, as well as the optical properties of these objects have been extensively studied (see for instance McDonald et al. 2014; Saro et al. 2015; Hennig et al. 2017; Chiu et al. 2018; Bulbul et al. 2019; Capasso et al. 2019a, and references therein). Furthermore, successful cosmological studies have been performed with the $\xi > 5$ subsample (Bocquet et al. 2015; de Haan et al. 2016; Bocquet et al. 2019), indicating that the survey selection function is well understood and that the mass information derived from the SZE is reliable. This motivates us to employ this sample as a reference for our validation of the observable mass relations and the selection function of the MARD-Y3 sample.

4.1.3 Cross-matched sample

To identify clusters selected both by SPT-SZ and by MARD-Y3, we perform a positional matching within the angular scale of 2 Mpc at the MARD-Y3 cluster redshift. We match 120 clusters in the redshift range $z \in (0.2, 1.1)$. We identify three clusters where the redshift determined by the MCMF run on RASS, $z_{\rm RASS}$, is significantly different from the redshift MCMF assigns for the SPT-SZ candidate, $z_{\rm SPT}$. Specifically, for these objects $|z_{\rm RASS} - z_{\rm SPT}| > 0.02(z_{\rm RASS} + z_{\rm SPT})/2$, which is equivalent to more than 3σ with respect to the typical MCMF photometric redshift accuracy (Klein et al. 2018, 2019). While for all three cases $z_{\rm RASS} < z_{\rm SPT}$, in all cases the MCMF run on the SPT-SZ candidate list identifies optical structures at $z_{\rm RASS}$ as well. Both their X-ray fluxes and SZE signals are likely biased with respect to the nominal relation for individual clusters due to the presence of several structures along the line of sight. Disentangling the respective contributions of the different structures along the line of sight is complicated by different scaling of X-ray flux and SZE signal with distance. We exclude these objects from the matched sample.

In only one case, two MARD-Y3 clusters are associated with the same SPT-SZ cluster: 'SPT-CL J2358-6129', $z_{\rm SPT} = 0.403$. Visual inspection (cf. Fig. B1) reveals that one of the MARD-Y3 clusters, $z_{\rm RASS} = 0.398$, is well centred on the SZE signal, and also coincides with a peak in the galaxy density distribution. The second MARD-Y3 cluster in the north–north-west, $z_{\rm RASS} = 0.405$, is offset from the peaks in galaxy density, and does not correspond to any SZE signal. Given the lack of the SZ-counterpart, we do consider this MARD-Y3 cluster not being matched by SPT. We also identify a pair of SPT clusters ('SPT-CL J2331-5051' and 'SPT-CL J2332-5053') matched to the same MARD-Y3 cluster (2RXS-J233146.5-505227), shown in Fig. B2. Both SPT clusters are at redshift ∼0.57, as is the MARD-Y3 clusters. The X-ray emission is blended into one source in the RASS image, but *Chandra* follow-up by (Andersson et al. 2011) clearly shows that 95 per cent of the X-ray originates from 'SPT-CL J2331-5051', which is also more significant in the SZe. We therefore take that to be the match.

4.2 Priors

In this section, we present the priors used in the likelihood analysis. We first discuss the cosmological priors assumed. Then, we describe the priors on the SZE signal–mass relation, the X-ray luminosity–mass relation, and the richness–mass relation. These priors are summarized in Table 1. In the respective subsections, we describe in which analysis step the specific prior is used.

Table 1. Summary of the informative priors employed in this work. These priors are implemented as Gaussian probability distributions, where we present the mean μ and the standard deviation σ as $\mu \pm \sigma$.

Cosmological parameters		
H_0	70.6 ± 2.6	Rigault et al. (2018)
Ω_M	0.276 ± 0.047	SPT (Bo19)
$S_8 = \sigma_8(\Omega_M/0.3)^{0.2}$	0.766 ± 0.025	SPT (Bo19)
SZE ζ–mass relation		
A_{SZ}	5.24 ± 0.85	SPT (Bo19)
B_{SZ}	1.53 ± 0.10	
C_{SZ}	0.47 ± 0.41	
σ_{SZ}	0.16 ± 0.08	
X-ray L_X–mass relation		
A_X	4.20 ± 0.91	SPT–XMM(Bu19)
B_X	1.89 ± 0.18	
C_X	-0.20 ± 0.50	
σ_X	0.27 ± 0.10	
Optical λ–mass relation		
A_λ	71.9 ± 6.1	SPT-DES (S15)
B_λ	1.14 ± 0.20	
C_λ	0.73 ± 0.76	
σ_λ	0.15 ± 0.08	

4.2.1 *Priors on cosmology*

Throughout this work, we marginalize over the following cosmological parameters to propagate our uncertainty on these parameters. The X-ray flux–mass relation has a distance dependence making it dependent on the present-day expansion rate, also called the Hubble constant H_0. We therefore adopt the prior $H_0 = 70.6 \pm 2.6$ km s^{-1} Mpc^{-1} from cepheid calibrated distance ladder measurements presented by Rigault et al. (2018).[8]

For our number counts analysis in Section 5.2, we constrain scaling relation parameters by comparing the measured cluster number counts to a prediction based on our scaling relation model with assumed cosmological priors. We assume priors $\Omega_M = 0.276 \pm 0.047$ and $S_8 = \sigma_8(\Omega_M/0.3)^{0.2} = 0.766 \pm 0.025$, derived by Bocquet et al. (2019, hereafter Bo19) from the number counts analysis of 343 SZE-selected galaxy clusters supplemented with gas mass measurements for 89 clusters and weak lensing shear profile measurement for 32 clusters. Note that the aforementioned H_0 prior is consistent with the constraints from Bo19. Furthermore, we assume that the Universe is flat and that the dark energy can be represented by a cosmological constant.

4.2.2 *Priors on SZE ζ–mass relation*

When performing the SPT-SZ cross-calibration (Section 5.1) we assume priors on the SZE scaling relation parameters to infer the X-ray flux–mass and richness–mass scaling relation parameters. These priors are derived from the X-ray and WL calibrated number counts of SPT-SZ selected clusters as described in Bo19. The adopted values are reported in Table 1. These priors were derived simultaneously with the cosmological priors discussed above, and both rely on the assumption that the SPT-SZ selection function is well characterized and that the SZE-signal–mass relation is well described by equation (4). These priors are also used when estimating the outlier fraction, the MARD-Y3 contamination and the SPT incompleteness (Section 5.4). Note that Bo19 only considered SPT-SZ clusters with SZE–signal-to-noise $\xi > 5$ and $z > 0.25$, while we adopt their results to characterize a sample with $\xi > 4.5$ and $z > 0.2$. Considering that this is an extrapolation from typical masses of $\sim 3.6 \times 10^{14}$ M$_\odot$ for $\xi = 5$ to $\sim 3.3 \times 10^{14}$ M$_\odot$ for $\xi = 4.5$, we view this as a minor change.

4.2.3 *Priors on X-ray* L_X*–mass relation*

The X-ray luminosity–mass relation (cf. Table 1) used as comparison for the luminosity–mass relations we derive from the SPT-SZ cross-calibration (Section 5.1) and the number count fits (Section 5.2) has been determined by Bu19, who studied the X-ray luminosities of 59 SPT-SZ-selected clusters observed with XXM-Newton.[9] The authors then use priors on the SZE-signal–mass relation to infer the luminosity–mass relation parameters. These measurements are also used as priors for the optical number counts (Section 5.2.1) and when determining the systematic uncertainty on the outlier probability π_t, the MARD-Y3 contamination π_c and the SPT-SZ incompleteness π_i (Section 5.4).

4.2.4 *Priors on optical λ–mass scaling relation*

The richness–mass relation used as comparison for the richness–mass relations we derive from the SPT-SZ cross-calibration (Section 5.1) and the number count fits (Section 5.2) was derived by S15 from a sample of 25 SPT-SZ-selected cluster, matched with DES redmapper selected clusters. In that work, the SZE-signal–mass relation parameters were determined by fitting the SPT-SZ-selected cluster number counts at fixed cosmology. The resulting constraints on the richness–mass relation are reported in Table 1. These measurements are also used as prior for the X-ray number counts (Section 5.2.2) and when determining the systematic uncertainty on the outlier probability π_t, the MARD-Y3 contamination π_c, and the SPT-SZ incompleteness π_i (Section 5.4).

5 APPLICATION TO MARD-Y3 AND SPT-SZ

In this section, we present the results of validation tests on the MARD-Y3 sample by way of examining the consistency of the X-ray–mass and the richness–mass scaling relations derived using different methods. First, we present the cross-calibration of the fluxes and richnesses using the externally calibrated SPT-SZ sample (Section 5.1). Then in Section 5.2, we derive the parameters of the X-ray–mass scaling relation from the X-ray number counts, the parameters of the richness–mass scaling relation from the optical number counts, and then explore the constraints on both scaling relations from a joint 2D X-ray and optical number counts analysis. We explore the implied cluster masses in Section 5.3, and in Section 5.4 we validate our selection functions by computing the probabilities of each cluster in one sample having a counterpart in the other and comparing these probabilities to the actual set of matched pairs and unmatched single clusters in each sample. This last exercise allows us to study outliers in observables beyond the measured scaling relation and observational scatter and has implications for the incompleteness in the SPT-SZ sample and the contamination in the MARD-Y3 sample.

[8]Given the still unresolved controversy on the exact value of the Hubble constant, the value adopted here has the benefit of not being in significant tension with any other published result.

[9]We use the relation of type II for the core included luminosity within the [0.5,2] keV band.

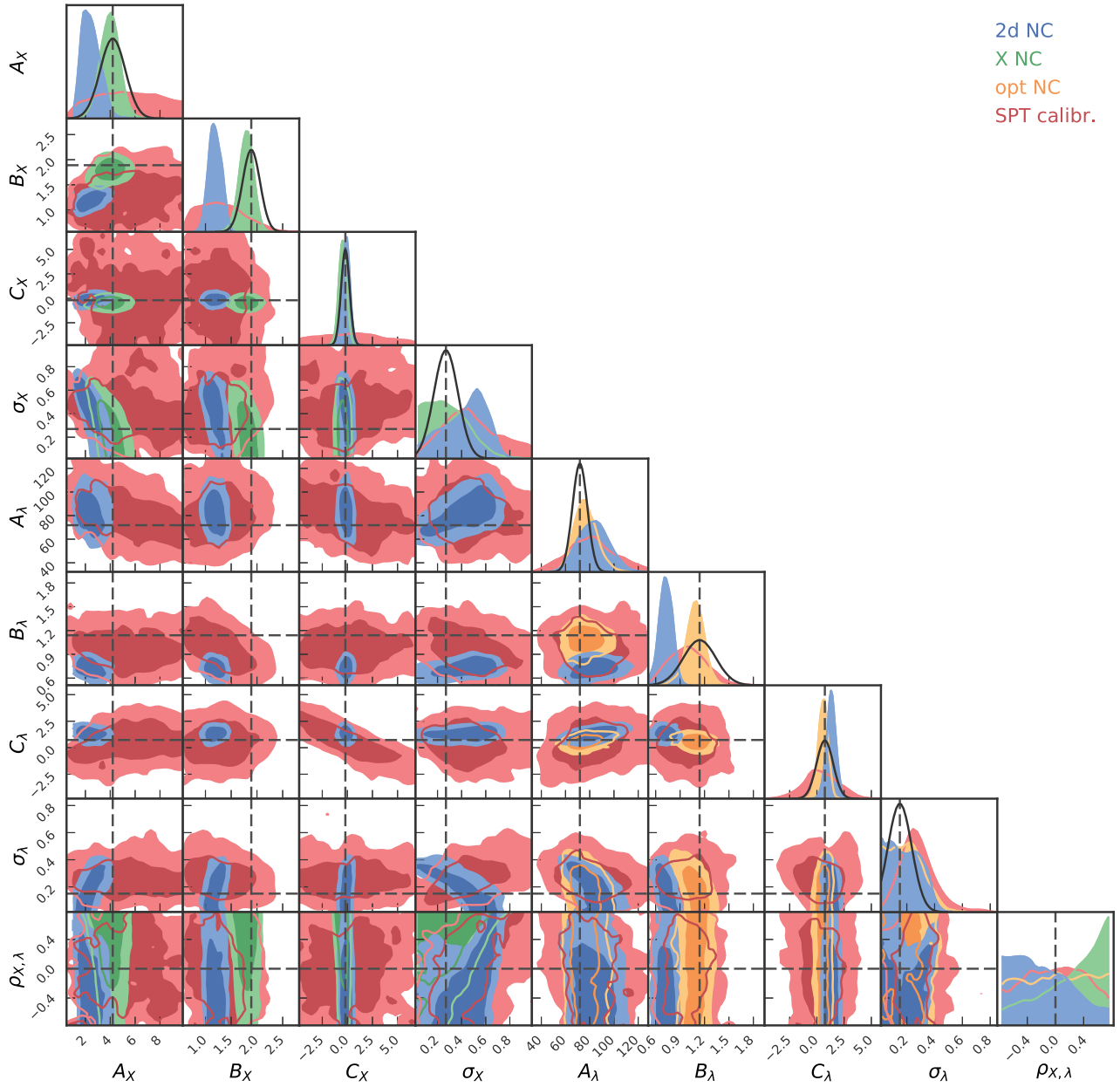

Figure 7. Marginal posterior contours of the free parameters in the SPT-SZ cross-calibration (SPT calibration, red), the number counts in X-ray flux and redshift (X NC, green), the number counts in richness and redshift (opt NC, orange), and the number counts in X-ray flux, richness, and redshift (2D NC, blue). In black, the literature values from Bu19 and S15. The SPT-SZ cross-calibration shows good agreement with the different number counts constraints and the literature. With the exception of the mass slope of the X-ray flux–mass relation inferred from 2D number counts, the constraints from the different number count experiments also show good agreement with the literature values. This provides strong evidence that our selection function modelling is adequate.

5.1 Validation using SPT-SZ cross-calibration

As implied in the methods discussion in Section 3.1, the results of the SPT-SZ cross-calibration of the MARD-Y3 mass indicators X-ray flux and richness are extracted by sampling the likelihood in equation (23). The free parameters of this fit are the parameters of the X-ray scaling relation (A_X, B_X, C_X, σ_X), of the richness scaling relation (A_λ, B_λ, C_λ, σ_λ) and the correlation coefficients between the intrinsic scatters ($\rho_{X,\lambda}$, $\rho_{X,SZ}$, $\rho_{\lambda,SZ}$). We put uniform priors on these parameters, whose bounds are adjusted to not cut off the posterior (such priors are hereafter referred to as 'wide, flat priors'). In the case of the correlation coefficients, we furthermore demand that $\det \mathbf{C} > 0$. We put informative priors on the parameters of the SZE-signal–mass relation (A_{SZ}, B_{SZ}, C_{SZ}, σ_{SZ}) and on the cosmological parameters (H_0, Ω_M, σ_8), as described in Section 4.2.

The resulting marginal posterior contours on the parameters without priors are shown in red (SPT calibration) in Fig. 7 and Table 2. The same figure also shows as a black line the literature values for these parameters, where we use Bu19 for the X-ray parameters, and S15 for the optical parameters. Our constraints are in agreement with these works, but display comparable or larger uncertainties despite the larger number of objects. This is due to different effects.

Table 2. Mean and standard deviation estimated from the 1D marginal posterior plots for the parameters of the X-ray and the richness scaling relation. Besides the constraints of the mass trend of the X-ray–mass relation and the corresponding intrinsic scatter, we find good agreement among our different analysis methods and with the literature values. This provides strong evidence that our selection function modelling is adequate.

	A_X	B_X	C_X	σ_X	A_λ	B_λ	C_λ	σ_λ
Literature	4.20 ± 0.91	1.89 ± 0.18	-0.20 ± 0.50	0.27 ± 0.10	71.9 ± 6.1	1.14 ± 0.20	0.73 ± 0.76	0.15 ± 0.08
SPT calibration	5.42 ± 2.48	1.31 ± 0.43	–	0.48 ± 0.23	81.6 ± 19.3	1.00 ± 0.22	0.39 ± 1.55	0.28 ± 0.13
X NC	3.97 ± 0.75	1.79 ± 0.14	-0.46 ± 0.38	0.28 ± 0.17				
opt NC					76.5 ± 9.3	1.09 ± 0.11	0.57 ± 0.44	0.20 ± 0.12
2D NC	2.45 ± 0.71	1.19 ± 0.12	-0.13 ± 0.37	0.42 ± 0.17	83.1 ± 12.3	0.72 ± 0.08	1.31 ± 0.43	0.19 ± 0.11

The difference between the sizes of the uncertainties on the richness–mass relation in this work and S15 are mainly due to the tighter priors on the SZE-signal–mass relation parameters utilized by S15. For instance, in S15 the prior on the amplitude of the SZE-signal–mass relation is four times smaller than the one used in this work. That being said, we here analyse a 4 times larger sample, which warrants at best an improvement of the constraints by a factor of 2. Our larger uncertainties on the richness–mass relation parameters are thus reflecting our more conservative treatment of systematic uncertainties on the SZE inferred masses.

This does not, however, explain why our constraints on the luminosity–mass relation are weaker than those reported by Bu19, as that work used priors on the SZE-signal–mass relation comparable to ours. Two different effects play a role in this case. (1) The measurement uncertainty on the luminosities extracted from pointed *XMM* observations is much smaller than on RASS-based luminosities. (2) Marginalizing over the systematic uncertainty on the matter density Ω_M and the Hubble parameter H_0 leads via the cosmological dependence of the luminosity distance and $E(z)$ to a systematic uncertainty $\delta C_X \sim 0.37$. This source of uncertainty is not considered in Bu19. In summary, our data are considerably less constraining then the *XMM* measurements, which themselves were analysed ignoring an important systematic uncertainty.

In Fig. 8 are plotted in different redshift bins the scaling relation between the intrinsic X-ray flux inferred from the X-ray flux error model (equation 10) and the intrinsic SZE signal-to-noise ratio inferred from the SZE error model (equation 13), as black points with 1σ and 2σ uncertainties. We also plot the predicted X-ray flux–SZE-signal relation obtained by combining the respective scaling relations. We show (black and grey) their marginalization over the Bo19 cosmological parameter and SZE-scaling parameter priors, the Bu19 X-ray-scaling parameter priors, and over the posterior of the SPT-SZ cross-calibration (red). As already noted from the contour plots of the marginal posteriors, our inferred scaling relation parameters are statistically consistent with the literature. Our calibration, however, prefers a steeper relation, which manifests also in the lower inferred value on the X-ray mass trend B_X.

The results for the SPT-SZ cross-calibration of the richness–mass relation are shown in Fig. 9. In different redshift bins, we plot as black point the intrinsic richness λ and the intrinsic SZE-signal ζ inferred from the respective error models (equations 12 and 13). We also plot the richness–SZE scaling derived from combining the richness–mass and the SZE-signal–mass relation. The resulting relation is shown with the uncertainties derived from the literature priors and the cross-calibration posteriors. The two constraints are in very good agreement. Yet, at high redshift $z > 0.5$, we note the presence of a high richness, low SZE-signal population, not well described by either the relation in the literature or our cross-calibrated relation. These objects will be discussed in more detail in Section 5.4.2.

5.2 Validation using number counts

As described in the method discussion in Section 3.2, we perform three different number counts experiments in this work: (1) we infer the X-ray flux–mass relation by fitting for the number counts of cluster as a function of measured flux and redshift; (2) we constrain the richness–mass relation by fitting for the number counts as a function of measured richness and redshift; and (3) we determine both relations by fitting the number of objects as a function measured flux, measured richness, and redshift.

5.2.1 X-ray number counts

While sampling the likelihood of the number counts in X-ray flux (equation 24), we let the parameters of the X-ray flux–mass relation (A_X, B_X, C_X, σ_X) float within wide, flat priors. We adopt priors on the relevant cosmological parameters (H_0, Ω_M, σ_8) as described in Table 1. We also put priors on the richness–mass relation parameter (A_λ, B_λ, C_λ, σ_λ). Furthermore, we empirically constrain the relation between X-ray detection significance ξ_X, measured flux $\hat{f}_X$, and exposure time t_{exp} from the sample. As described in more detail in Sections 2.3.1 and 6.1, this results in four tightly constrained nuisance parameters that impact the X-ray selection function. The resulting posteriors on the X-ray scaling relation parameters are shown in green in Fig. 7. We find tight agreement with the literature values, at comparable accuracy on the marginal uncertainties.

In Fig. 10, we plot the number counts in measured X-ray flux bins in three different redshift bins with the respective Poissonian errors. We also plot the prediction for the number of objects in the same bins, once marginalized over the literature values (black and grey), over our 1D fit (green), and our 2D number counts fit (blue, cf. Section 5.2.3). The 1D fit provides an accurate fit to the data, with the exception of the regime where X-ray incompleteness sets in, where it tends to slightly underestimate the number of clusters. The prediction from the literature provides a statistically consistent description of the data, albeit systematically more than 1σ, and less than 2σ low at low mass. These trends, while not statistically significant, are confirmed by inspecting the inferred masses from our posterior (see below Section 5.3).

5.2.2 Optical number counts

Just as the number counts as a function of measured flux can be used to infer the X-ray scaling relation parameters, the number counts in richness can be used to infer the richness–mass relation parameters. To this end, we sample the likelihood of number counts in richness bins (equation 27). We let the parameters of the richness–mass relation (A_λ, B_λ, C_λ, σ_λ) free within wide, flat priors, while we adopt informative priors from the literature on the cosmological parameters (H_0, Ω_M, σ_8). Importantly, modelling the X-ray incompleteness in

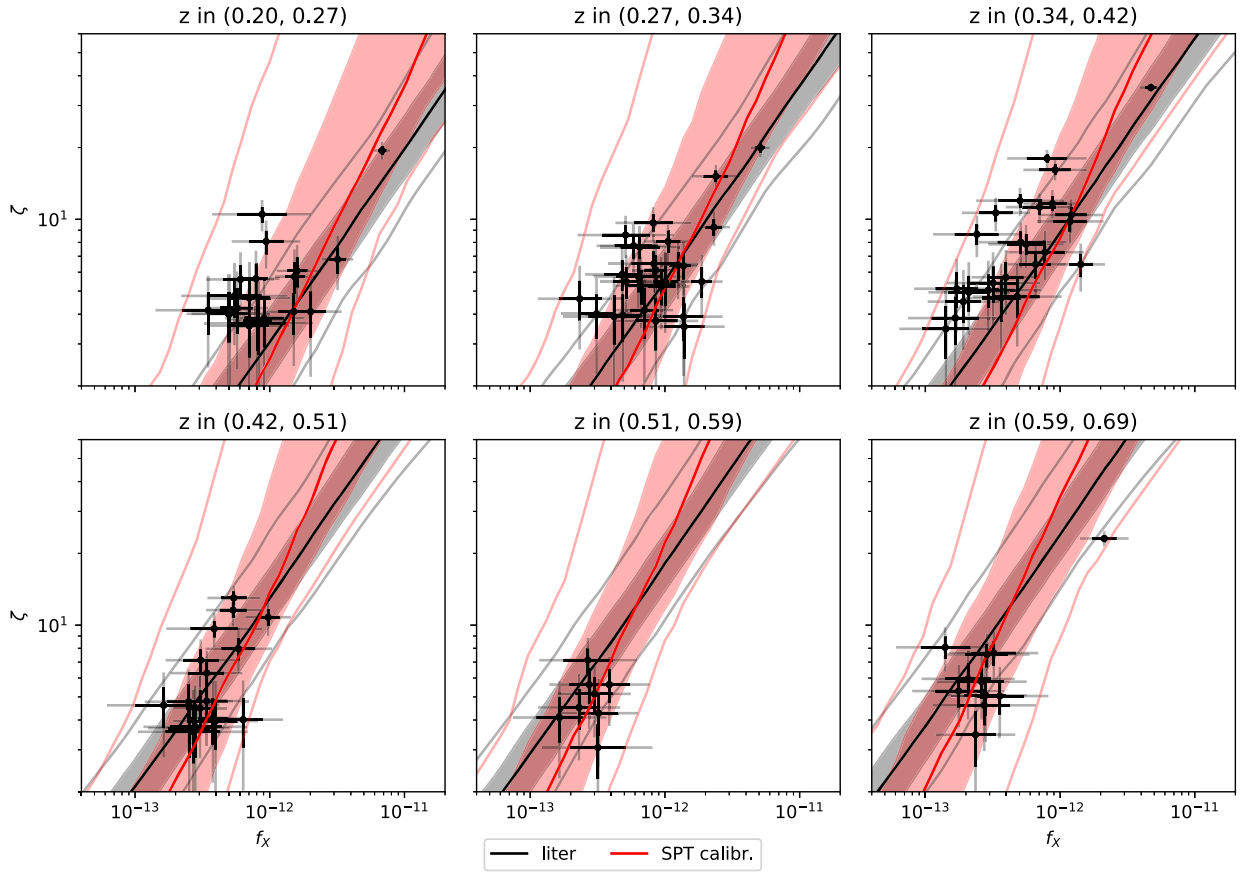

Figure 8. Black points mark the intrinsic X-ray flux f_X and SZE–signal to noise ζ inferred from the respective error models for our cross-matched sample. The X-ray flux–SZE-signal relation is shown either marginalized over the literature priors (black and grey) or over the posterior of our cross-calibration to SPT-SZ (red). The full lines are the median values, the filled region covers the range from the 16th to 84th percentile and the transparent lines show the 2.5th and 97.5th percentile. While both sets of scaling relation parameters are consistent, we find a tendency for a weaker mass trend in the X-ray observable than reported in the literature.

the space on measured richness requires a way to transform from measured richness to X-ray flux. Thus, while the transformation from richness to mass is fit, we need to assume a transformation from mass to X-ray flux. This is done by putting priors on the X-ray scaling relation parameters (A_X, B_X, C_X, σ_X). As for the X-ray number counts, we empirically constrain the relation between X-ray detection significance ξ_X, measured flux $\hat{f}_X$, and exposure time t_{exp} from the sample and predict the X-ray selection function on the fly.

The resulting marginal posterior contours are shown in Fig. 7 in orange. We find good agreement with the literature values and with the SPT-SZ cross-calibration. The marginal uncertainties are comparable to the literature values, despite being marginalized over cosmological parameters. We also find that the constraints from the number counts are more stringent than those derived from the SPT-SZ cross-calibration.

One can visually assess the quality of the resulting fit in Fig. 11, where we plot the number of objects in measured richness for different redshift bins as black points with Poissonian error bars. We also plot the predicted number of objects with the uncertainties derived from the literature priors (black and grey) from our 1D fit (orange), and our 2D number counts fit (blue, cf. Section 5.2.3). Also in this case we note that the literature prediction is systematically between 1σ and 2σ low, which manifests also in different mass estimates (see below in Section 5.3).

5.2.3 *Combined X-ray and optical number counts*

We also fit for the abundance of clusters as a function of measured X-ray flux $\hat{f}_X$, measured richness $\hat{\lambda}$, and redshift, which we will refer to a '2D number counts', by sampling the likelihood in equation (30). We allow the parameters of both the X-ray scaling relation (A_X, B_X, C_X, σ_X) and the richness scaling relation (A_λ, B_λ, C_λ, σ_λ) to float within wide, flat priors. We adopt priors on the cosmological parameters from Table 1. Furthermore, we empirically constrain the relation between X-ray detection significance ξ_X, measured flux $\hat{f}_X$, and exposure time t_{exp} from the MARD-Y3 sample and predict the X-ray selection function on the fly (cf. Sections 2.3.1 and 6.1).

In Fig. 7, we show the marginal posterior contours on the scaling relation parameters in blue. We find good agreement with the results from the SPT-SZ cross-calibration on all parameters. When comparing the constraints from 2D number counts (blue) on the X-ray scaling relation parameters to the constraints from the number counts in X-ray flux (green), we find good agreement on the values of the amplitude and redshift evolution. However, we find a shallower X-ray observable mass trend than from the X-ray number counts, and we see a similar shift in the optical mass trend parameter, although in this case the statistical significance is small. Given the agreement of the X-ray number counts result is with Bu19, the results from the 2D number counts are in some tension with both. As show in

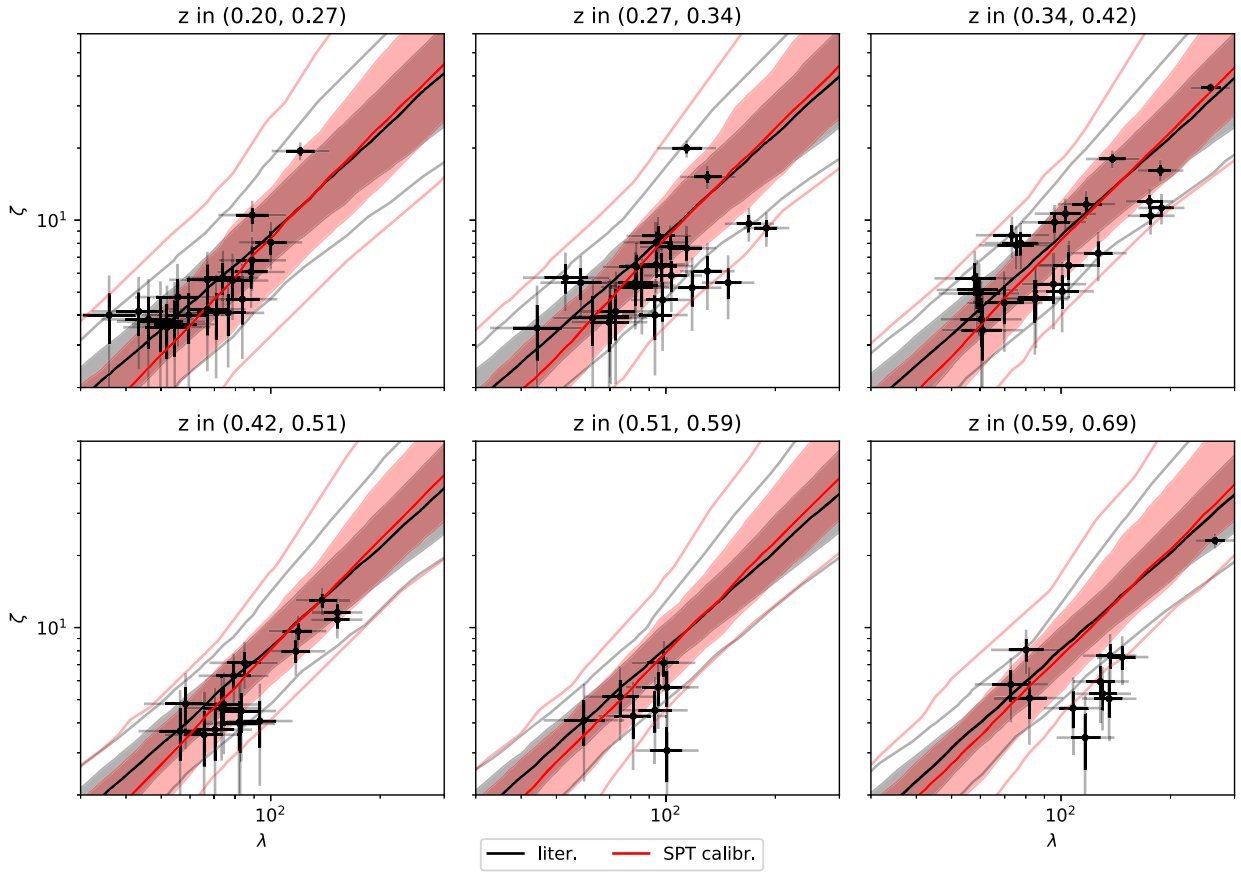

Figure 9. Black points mark the the intrinsic richness λ and SZE-signal-to-noise ζ inferred from the respective error models for our cross-matched sample. The richness–SZE-signal relation is shown either marginalized over the literature priors (black) or over the posterior of our cross-calibration to SPT-SZ (red). The full lines are the median values, the filled region covers the range from the 16th to 84th percentile and the transparent lines show the 2.5th and 97.5th percentile. Both sets of scaling relation parameters are consistent, at $0.51 < z < 0.69$, however, they fail to describe a part of the population with low SZE signal and high richness.

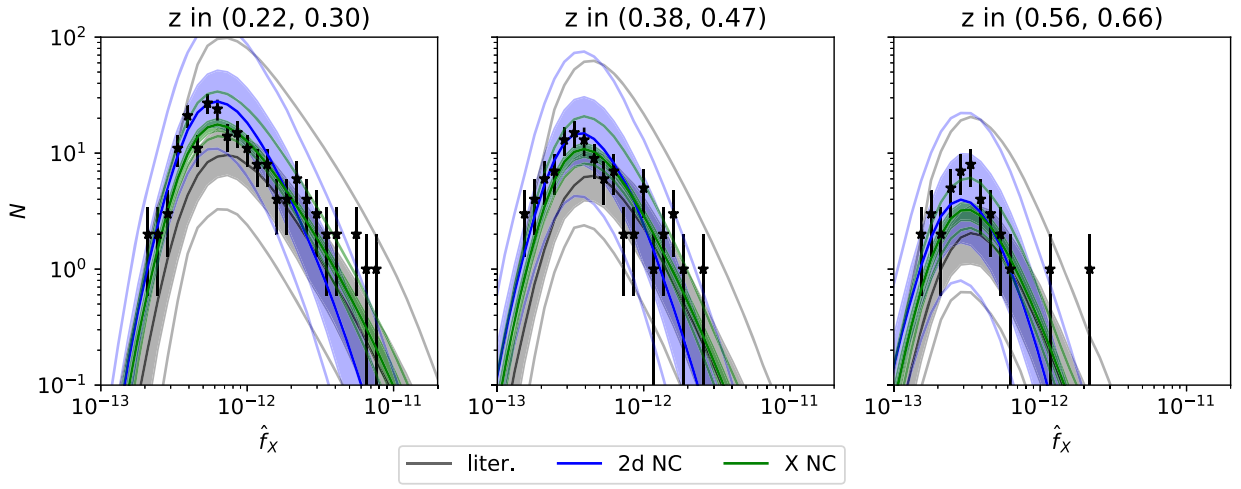

Figure 10. Measured number of clusters in bins of measured X-ray flux for different redshift bins as black points with Poissonian error bars. We overplot the prediction for the number of objects with the uncertainties derived from the literature values (black and grey), from our 1D fit (green), and from our 2D fit (blue). The full lines are the median values, the filled region covers the range from the 16th to 84th percentile and the transparent lines show the 2.5th and 97.5th percentile. The latter captures adequately both the increasing rarity of high-redshift objects, as well as the effect of X-ray incompleteness at low flux. The measurement is also consistent with the literature values.

Section 5.3 below, these constraints however do not results in statistically inconsistent mass estimates. Nevertheless, possible systematic effects impacting our validation tests are discussed in Sections 6.1 and 6.2.

Of interest is also the constraint the 2D number counts put on the two intrinsic scatters in X-ray flux and richness. Inspecting their joint marginal posterior in Fig. 7 reveals a distinct degeneracy in the form of an arc. This is the natural result of the fact that the

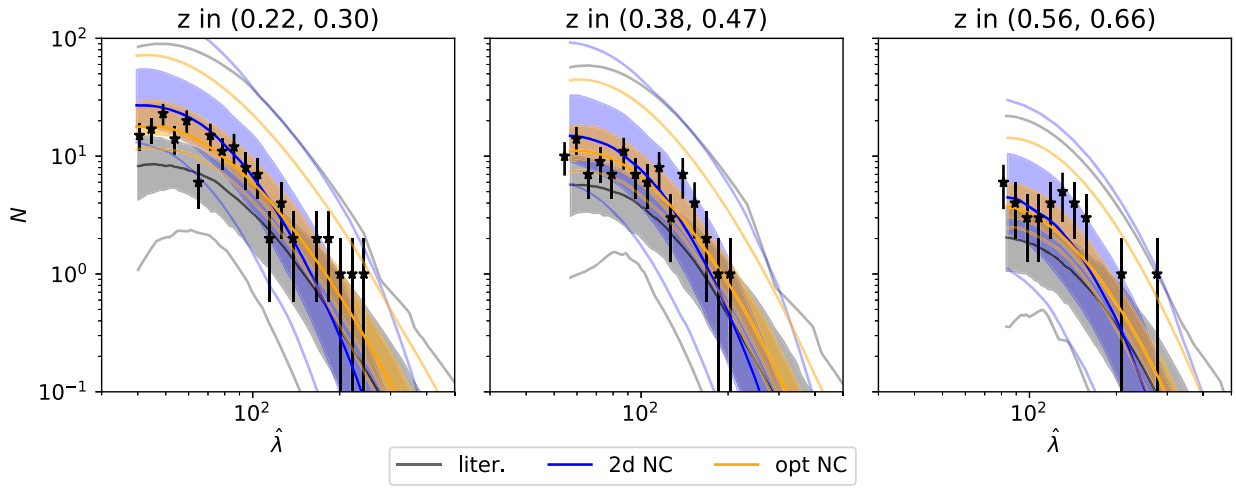

Figure 11. Number of objects in bins of measured richness $\hat{\lambda}$ in different redshift bins as black points with Poissonian error bars. Overplotted is the expected number of objects as a function of measured richness for the same bins marginalized over the uncertainties from the literature (black and grey), from our 1D fit (orange), and from our 2D fit (blue). The full lines are the median values, the filled region covers the range from the 16th to 84th percentile and the transparent lines show the 2.5th and 97.5th percentile. The shape of the abundance at low richness would in principle be closer to a power law, but at low richness, the X-ray selection of the samples leads to a decrease in the number of objects, which is well fitted by our selection model.

2D number counts can only constraint the total scatter between the two observables, but not the two individual scatters between each observable and mass. The total scatter between observables, being the squared sum of the individual scatter, sets the radius of the arc. Noticeably, this arc-like degeneracy excludes the possibility that both the X-ray and the richness scatter are small.

For visual inspection of the 2D number counts fit in Fig. 12 we present the distribution in measured X-ray flux and measured richness of our sample in different redshift bins as black stars. We also plot the contours of the predicted number of objects in equally spaced logarithmic bins (shown by the overlaid grid): in blue, the prediction for the best-fitting value of the 2D number counts, while in grey the prediction from the literature. The selection in richness due to the $f_{\rm cont} < 0.05$ cut is at every redshift a sharp cut in measured richness, as can be seen up to the intra bin scatter due to the large bins used for plotting. The effect of the X-ray selection function is harder to see, but can be appreciated in the shape of the contours at low flux: they show a bend, predicting very small numbers of objects at the lowest fluxes. Notably, the distribution of the data displays a large dispersion, which is better captured by our fit (blue) than by the prediction from the literature (grey). This confirms that the measurement of a larger X-ray scatter is indeed a feature of the data visible in the 2D cluster abundance. Despite the larger intrinsic scatter, 2D number counts posterior provide also a prediction of the X-ray and optical 1D number counts that is consistent with the data within the systematic and statistical uncertainties, as can be see by the blue predictions in Figs 10 and 11.

5.3 Validation using cluster masses

In this section, we investigate the prediction of the individual halo masses derived from the different constraints on the scaling relation parameters extracted above. Given that the number of objects as a function of mass is known, this section quantifies the relative goodness of fit of the number counts between the different fits we performed (X-ray, optical, and combined).

To estimate the masses for each cluster given its measured X-ray flux $\hat{f}_{\rm X}^{(i)}$ (or analogously the measured richness $\hat{\lambda}^{(i)}$), we compute the distribution of probable masses

$$P(M|\hat{f}_{\rm X}^{(i)}, z^{(i)}, \boldsymbol{p}) \propto \int {\rm d}f_{\rm X} P\left(\hat{f}_{\rm X}^{(i)}|f_{\rm X}\right) P(f_{\rm X}|M, z^{(i)}, \boldsymbol{p}) \times \frac{{\rm d}N}{{\rm d}M}\Big|_{M,z^{(i)},\boldsymbol{p}}, \quad (37)$$

where $P(f_{\rm X}|M, z^{(i)})$ is the mapping between intrinsic flux and mass obtained by only considering the first component of equation (5). Note also that the above equation needs to be normalized in such a way that $\int {\rm d}M\ P(M|\hat{f}_{\rm X}^{(i)}, z^{(i)}, \vec{p}) = 1$, which sets the proportionality constant.

The X-ray mass $M_{\rm X}$ (and analogously the optical mass M_λ) can then be estimated to be

$$\ln M_{\rm X}^{(i)}|_{\boldsymbol{p}} = \int dM\, P\left(M|\hat{f}_{\rm X}^{(i)}, z^{(i)}, \boldsymbol{p}\right) \ln M. \quad (38)$$

Note that these masses naturally take account of the Eddington bias, which is fully described by equation (37).

The X-ray and optical masses are affected by systematic uncertainties in the scaling relation and cosmological parameters. We capture this uncertainty in each case by marginalizing the mass posterior over the appropriate posterior distribution of the parameters that we determined above. We marginalize the mass over different scaling relation parameter posteriors, including those from the literature (liter.), those from the SPT-SZ cross-calibration (SPT calibr.), and those from the combined X-ray and optical number counts (2D NC), the X-ray number counts (X NC) and the optical number counts (opt NC). The mass posteriors are derived for all clusters in the MARD-Y3 sample.

In the upper row of Fig. 13, we present the ratio between the X-ray masses derived from our posteriors to the X-ray masses obtained from the literature (Bu19) as a function of inferred literature mass (left-hand panel) and of redshift (right-hand panel). We find that the mass inferred from the number counts in X-ray flux is consistent with the literature values, while the masses inferred from the 2D number counts and the SPT-SZ calibration are lower than the literature masses. In the case of the SPT masses, the difference never exceeds

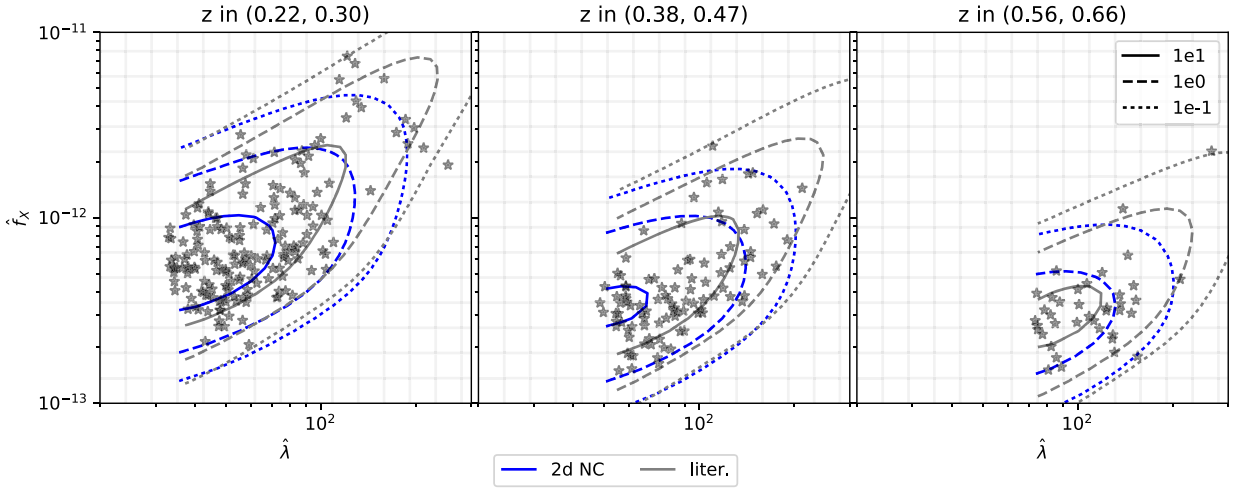

Figure 12. Number counts in both measured richness $\hat{\lambda}$ and measured X-ray flux $\hat{f}_X$ visualized by presenting in different redshift bins the distribution of our sample (points), the contours of the 2D predicted number of objects for the literature values (grey, liter.) and for the maximum-likelihood point of our fit to the data (blue, 2D NC). The number of objects shown in the contours (10, 1, 0.1 objects contours as full, dashed, and dotted lines respectively) refers to the bins shown in the overlaid grid. Our 2D fit prefers larger scatter and provides a better description of the data than does the literature prediction.

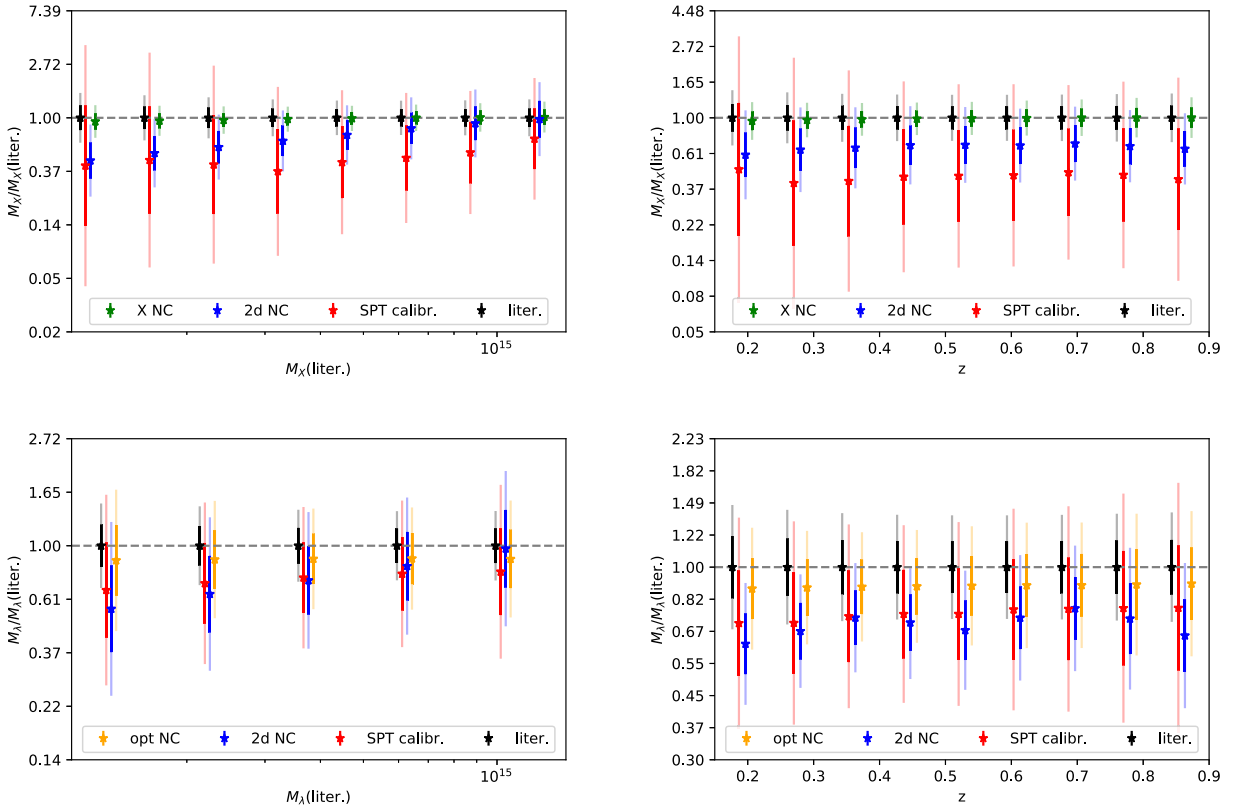

Figure 13. Ratio of the masses derived by our analysis methods and the masses derived from the literature values, for masses inferred from their measured X-ray flux (M_X, upper row) and their measured richness (M_λ, lower row), as functions of mass (left-hand column) and redshift (right-hand column), together with the 1σ and 2σ systematic uncertainties on the individual masses due to the incomplete knowledge of the scaling relation parameters. All masses refer to spherical over densities 500 times the critical density of the Universe. The masses we recover from SPT-SZ cross-calibration (SPT calibr., red) and different flavours of number counts, while being in most cases systematically low, are statistically consistent with the masses inferred by adopting the literature values. Tension beyond 1σ, but still smaller than 2σ appears at the low-mass end of the inferred X-ray masses.

1σ at all redshifts and masses we considered. For the 2D number count masses, we find that they are 1σ low at all redshifts, and up to 2σ low at masses of 1–2 $\times 10^{14}$ $M_\odot$. At masses of around 10^{15} $M_\odot$ they are in perfect agreement with the other mass estimates. This is due to the different values of inferred mass trend. As a function of redshift, the masses inferred from 2D number counts and the SPT-SZ calibration are also lower, reflecting on one side the prevalence of low-mass systems. On the other side, this shift is also due to the

larger intrinsic scatter recovered from the 2D number counts and the SPT-SZ calibration, that together with the shallower mass slope leads to a larger intrinsic mass scatter. This results in larger Eddington bias corrections and ultimately lower inferred masses. Note that the plotted error bars are strongly correlated (e.g. shifting A_X up by 1σ shifts all masses down by 1σ). At the current level of statistical and systematic uncertainty we conclude that different methods predict mutually consistent individual masses from the X-ray flux at less than 2σ. Yet the magnitude of the intrinsic scatter of the X-ray luminosity at fixed mass and redshift, together with its mass trend, are indications of possible internal tensions and unresolved systematics. These trends where already noted when comparing our best-fitting number count models to the data (see above in Section 5.2) and will be discussed further in Section 6.2.

In the lower row of Fig. 13, we also show the ratio between the optical mass inferred from our fits to the value taken from the literature (S15). Here, we find that all our methods provide a lower, yet statistically consistent mass estimate. The difference is likely due to an analysis choice in the literature values. Namely, S15 utilizes priors for the SZE–scaling relation parameters derived from fitting the SZE number counts at fixed cosmology. In that work, however, the CMB derived cosmology from Planck Collaboration (2014) was used, which results in $A_{\rm SZ,\,S15} = 4.02 \pm 0.16$, and therefore is an overestimation of masses by $\sim$18 per cent compared to our work. This shift accounts for most of the shifts seen in M_λ here. Even without this correction, at the current level of systematical uncertainties, the individual optical masses inferred from our different analysis methods are mutually consistent. This is expected because our $A_{\rm SZ}$ prior is consistent with the value used by S15. Furthermore, while 2D number counts predict a shallower mass trend than all other methods, in the mass range we consider this does not lead to significant tension with the other analysis methods.

This consistency check of mass estimates underscores the importance of weak-lensing mass calibration as a component of the validation of cluster samples. If the cosmology marginalized constraints on cluster masses from weak lensing are not consistent with those from cluster counts, then that would be clear evidence of an inadequacy in the selection model or an unaccounted for bias in the weak-lensing calibration analysis. As noted previously, we will examine the validation with the weak-lensing constraints in a forthcoming analysis.

5.4 Validation using independent cluster samples

Having established in the section above that our selection function modelling allows us to infer the masses of the MARD-Y3 clusters consistently within the systematic uncertainties, we now move on to a further test of the selection functions of the two samples.

As described in the methods Section 3.3, we investigate the SPT-SZ and MARD-Y3 selection functions by comparing the probability of each MARD-Y3 object being detected by SPT-SZ to the actual occurrence of such a detection. As established in Section 4.1.3, there are 120 clusters in the cross-matched sample, but the validation we do here also uses information from unmatched clusters. We first consider the MARD-Y3 sample and compute the SPT-SZ detection probability for each of these objects. Comparing these probabilities to the actual occurrence of matches provides an estimate of SPT-SZ incompleteness as well as MARD-Y3 contamination. We then consider the SPT-SZ sample and compute the probability that an SPT-SZ cluster is detected in MARD-Y3. In this case, we also constrain the outlier fraction beyond the lognormal scatter, more precisely the fraction of objects with an abnormally high X-ray flux and optical richness, or a surprisingly low SZE signal.

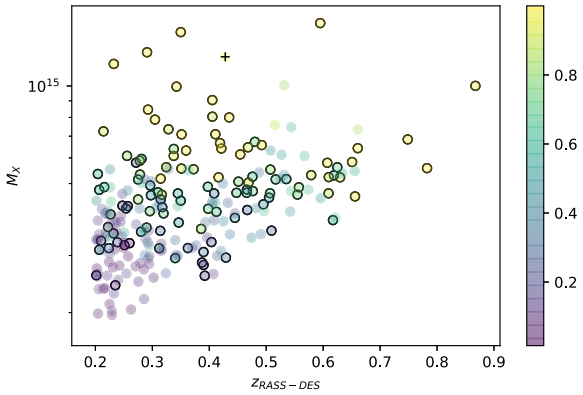

Figure 14. MARD-Y3 sample in the joint SPT-DES Y3 footprint at redshift $z > 0.2$. Colour encodes the probability of an SPT-SZ detection for each object, showing the characteristic mass selection of the SPT-SZ catalogue. Black circles indicate matched clusters, while the cross marks the missing SPT-SZ confirmations of MARD-Y3 objects. The only missed SPT confirmation is due to a catastrophic redshift error in the application of MCMF to the SPT-SZ sample.

5.4.1 SPT-SZ detection of MARD-Y3 clusters

In Fig. 14, we show the MARD-Y3 cluster sample in the joint SPT-DES Y1 footprint, plotted as a function of the X-ray-derived mass and the redshift presented by K19. Note that the mass used in this plot is used solely for presentation purposes, and does not go into any further calculation. We colour-code the MARD-Y3 clusters based on their SPT-SZ detection probability $p^{(i)}_{\rm SIM}$, computed following equation (35). This prediction reflects the mass information contained in each cluster's measured flux $\hat{f}^{(i)}_{\rm X}$ and measured richness $\hat{\lambda}^{(i)}$. It also nicely visualizes the approximate mass selection at $M \gtrapprox 3 \times 10^{14}\,{\rm M}_\odot$ of the SPT-SZ sample.

We place black circles around the matched clusters. When determining the detection probabilities using the literature values for the scaling relation parameters, we identify six clusters that have high detection probability, but are not matched, so-called missed SPT confirmations of MARD-Y3 objects. However, when determining the detection probabilities either from the posterior of our SPT-SZ cross-calibration or the 2D number counts, only one of these systems is identified as a missed confirmation: 2RXS J033045.2−522845.

This object coincides in the sky with the NE part of A3128. It has been found to be a $z \sim 0.43$ cluster by Werner et al. (2007) using *XMM* observations, by ACT observations (Hincks et al. 2010) and the SPT-SZ survey (Bleem et al. 2015), despite the large number of $z \sim 0.05$ galaxies in the foreground (also visible the DES image in the upper left panel of Fig. B5). The redshift $z \sim 0.43$ is confirmed by MARD-Y3. However, application of MCMF on the SPT-SZ sample found $z_{\rm SPT} = 0.056$, sourced by the foreground galaxies. Consequently, this object is erroneously not included in our SPT-SZ sample, which has the redshift selection $z_{\rm SPT} \in (0.2, 1.)$. Noticeably, MCMF run on SPT-SZ finds also a structure with $f_{\rm cont,SPT} = 0$ at $z \sim 0.43$. It is however discarded by the automated highest peak selection as also the $z \sim 0.05$ structure has $f_{\rm cont,SPT} = 0$. In summary, this object is missing in our SPT-SZ sample due to a catastrophic MCMF failure when run on SPT-SZ. To keep the pipeline automated and avoid human decision making, we do not apply any special treatment to this object.

We also aim to constrain the occurrence of contamination in the MARD-Y3 sample by introducing the probability π_c that an MARD-Y3 object is not a cluster, and should not therefore be detected by SPT. Simultaneously, we also introduce the SPT-SZ incompleteness π_i, the probability that any MARD-Y3 cluster that should have been detected in the SPT-SZ survey was not (cf. Fig. 5). This allows us to use the actual list of detections and non-detections together with the raw probabilities of detection to constrain these extra probabilities, as discussed in equation (36). We find that π_c and π_i are degenerate parameters, with only the difference between the two values constrained by our data, rather than the two values separately. Under the assumption of an MARD-Y3 contamination of $\pi_c = 0.025$, as derived by K19 for the $f_{\rm cont} < 0.05$ sample used here, we find $\pi_i = 0.284 \pm 0.043(\text{stat.})^{+0.108}_{-0.186}(\text{sys.})$, when marginalizing over the literature priors. When marginalizing over the SPT-SZ calibration posterior we find $\pi_i < 0.030(\text{stat.})$ and $\pi_i < 0.030(\text{sys.})$ at 68 per cent confidence, while we find $\pi_i < 0.047(\text{stat.})$ and $\pi_i < 0.231(\text{sys.})$ at 68 per cent confidence when marginalizing over the 2D number counts constraints together with the priors from Bo19 on the SZE–signal scaling relation.

The difference in inferred central value for the SPT-SZ incompleteness is due to the different mass predictions when using the literature priors as compared to our fits. As discussed in Section 5.3, our SPT-SZ cross-calibration and our 2D number counts analysis imply lower X-ray and optically derived masses than the literature priors. This systematically lowers the SPT-SZ detection probability of MARD-Y3 clusters, resulting in different incompleteness probabilities when comparing to the actual number of matched objects. We interpret this as another piece of evidence that the SPT-SZ cross-calibration and the 2D number counts provide a more accurate picture of the observable–mass relation than the literature priors. In fact, they reveal that the scatter around our luminosity–mass relation is larger than the scatter found by Bu19. Yet, within the statistical and systematic uncertainties the results are still in agreement.

Another interesting aspect is the magnitude of the statistical and systematic uncertainty on the SPT-SZ contamination. Note that the statistical uncertainties when marginalizing over the different posteriors are comparable. This reflects the fact that they are derived from a sample of a given size. The minor differences can be appreciated by noting that in equation (36) the individual clusters likelihood of π_i are weighted by the detection probabilities $p^{(i)}_{\rm SIM}$, which are different depending on which posterior is used to compute them. On the other hand, the magnitude of the systematic uncertainty introduced by the marginalization over the different posteriors is quite different. Marginalizing over the SPT-SZ cross-calibration posterior provides the smallest systematic uncertainty. This is expected when considering that the SPT-SZ cross-calibration constrains $P(\zeta|\hat{f}^{(i)}_{\rm X}, \hat{\lambda}^{(i)}, z^{(i)})$ (cf. equations 21–23), which is the major source of systematic uncertainty when computing the SPT-SZ detection probabilities of MARD-Y3 cluster (cf. equation 35). These distributions are predicted less accurately by the literature priors and the 2D number counts.

5.4.2 *MARD-Y3 detection of SPT-SZ clusters*

We also test the MARD-Y3 selection function by computing the probability of detecting each of the SPT-SZ clusters in the DES-Y3 footprint. In Fig. 15, we show the SPT-SZ sample as a function of redshift and SZE-derived mass. Note that the SZE-derived mass shown in this figure is only used for presentation purposes. Colour encodes the MARD-Y3 detection probability, computed via equation (33). The colour coding reflects the approximate flux selection of the MARD-Y3 sample. We highlight the matched clusters with black circles.

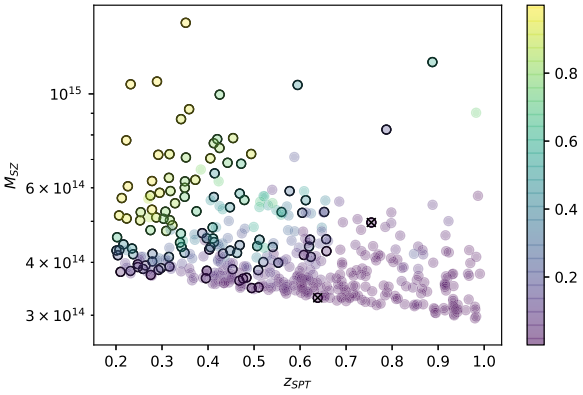

Figure 15. SPT-SZ selected sample in the joint SPT-DES Y3 footprint. Colour encodes the probability of MARD-Y3 detection, showing the characteristic flux selection of an X-ray survey. Black circles indicate matched clusters, while crosses mark unexpected MARD-Y3 confirmations of SPT-SZ objects (having an MARD-Y3 detection probability <0.025 but matched none the less).

Out of the 120 clusters in the cross-matched sample, we identify five unexpected MARD-Y3 confirmations, which are SPT-SZ clusters that show up in MARD-Y3 even though they have an MARD-Y3 detection probability <0.025, as calculated by marginalizing over the literature values. This list is expanded by three more unexpected confirmations when marginalizing over the 2D number counts posterior. When marginalizing over the SPT-SZ cross-calibration posterior, we find two unexpected confirmations, all of which are in common with the aforementioned ('SPT-CLJ0324-6236', visual inspection in Fig. B4, and 'SPT-CLJ0218-4233', visual inspection in Fig. B3). These clusters are marked in Fig. 15 with crosses. Visual inspection of this object reveals that in both cases a clear X-ray structure coincides with a significant SZE signal and a red sequence galaxy overdensity at the cluster redshift. Thus, these objects are likely genuine clusters with multiwavelength properties that are not captured by our scaling relation model and the scatter around it. These objects also do not display any exceptional behaviour with respect to the mean distribution in X-ray flux, richness.

Given these indications, we investigate the probability that an SPT-SZ object that should not be matched by MARD-Y3 is matched anyway. It cannot a priori be excluded that the distribution of X-ray luminosities or SZE signals at fixed mass has in actuality tails extending beyond the lognormal scatter model we assumed in Section 2.1. Such tails would lead to unexpected detections. The probability of a cluster living in such a tail, that is being an outlier, is given by the parameter π_t (see Section 3.3.1). Taking account of the detection probabilities and the actual occurrence of detections, we use the likelihood presented in equation (34). We find $\pi_t = 0.059 \pm 0.017(\text{stat.}) \pm 0.031(\text{sys.})$, when marginalizing over the literature priors. When marginalizing over the SPT-SZ cross-calibration posterior we find $\pi_t = 0.061 \pm 0.018(\text{stat.}) \pm 0.040(\text{sys.})$, while we find $\pi_t = 0.095 \pm 0.018(\text{stat.}) \pm 0.019(\text{sys.})$ when marginalizing over the 2D number counts constraints together with the priors from Bo19 on the SZE-signal scaling relation.

These constraints are mutually consistent in a statistical sense. Yet, the significance of the detection of the tail beyond lognormality ranges from 1.4σ for the literature priors, over 1.6σ for the SPT-SZ cross-calibration to 3.6σ for the 2D number counts constraints.

Internal inconsistencies in the number counts (discussed in Sections 5.2, 5.3, 6.1, and 6.2) might affect the latter result. Better mass information is required to distinguish whether our findings are a statistical fluke, the result of an unresolved systematic or stem from a genuine signal. If the presence of a lognormal tail would be confirmed, more detailed observations are needed to understand the source of the outliers we selected. For example, high-angular resolution X-ray or mm-wave observations, in combination with the spectroscopic optical data, would help to rule out any astrophysical confusion in either the X-ray or SZ measurements and identify any lower mass structures or objects along the line of sight, which could be affecting any of the observables.

6 DISCUSSION

Here, we first summarize the findings from the previous section and then discuss implications. We focus on different aspects, including: (1) internal indications for unresolved systematics in the selection function modelling, (2) the outcome of our validation, (3) the impact of optical incompleteness and the resulting benefits from its modelling, and finally (4) the implications of this work for cosmological studies.

6.1 X-ray selection function systematics

In Section 2.3.2, we discussed potential unresolved redshift trends of the selection function fit. Given that the X-ray selection spans a mass range of factor of 3 (see for instance Fig. 6) from low to high redshift, residual redshift trends in the X-ray selection functions are likely to impact the inferred mass trend as much as they are likely to impact the redshift trend of the X-ray flux–mass relation. This systematic manifests itself in different places, as discussed in the following.

When sampling the X-ray number counts (cf. Section 5.2.1), we sample the parameters of the richness–mass scaling relation with priors from the literature to estimate the effect of optical incompleteness of the sample. While the prior on the redshift evolution is $C_\lambda = 0.73 \pm 0.76$, the posterior is $C_\lambda = 0.34 \pm 0.53$, indicating that the X-ray number counts likelihoods slightly prefer a weaker redshift trend of the richness, effectively making the optical incompleteness larger at low redshift than at high redshift. This preference may be compensation for the fact that our model seems to predict small an X-ray selection function at high redshift.

Similarly, when sampling the optical number counts, we rely on priors on the X-ray flux–mass scaling relation to propagate the X-ray selection function to the space of measured richness. Also in this case the prior $C_X = -0.20 \pm 0.50$ is altered to a posterior $C_X = -0.50 \pm 0.38$. Consequently, a weaker redshift trend is preferred by the number counts, possibly as in an attempt to compensate the same residual systematic effect. Lastly, we find that the X-ray, as well as the optical number counts, pull the prior we placed on $\Omega_M = 0.276 \pm 0.047$ to a posterior $\Omega_M = 0.296 \pm 0.038$ from X-ray number count, and $\Omega_M = 0.302 \pm 0.037$ from optical number counts, respectively. If these shifts result in biases of the cosmological results once direct mass information from weak lensing is available, they should be further investigated.

As described in the case of a putative redshift residual, the empirical calibration of the selection function provides an opportunity to uncover unresolved systematics. From this perspective it offers advantages in comparison to selection functions determined from image simulations. For instance, consider in Fig. 16, the posterior constraints on the significance–flux scaling parameters resulting from fitting either directly to the relevant catalogue data by sampling

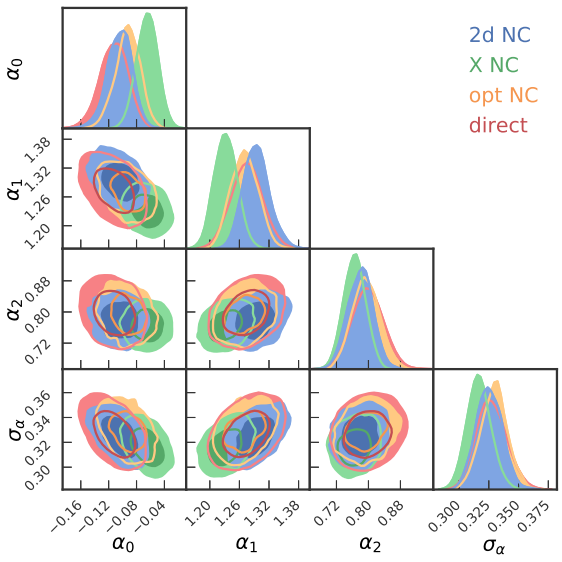

Figure 16. Marginal posterior contours on the extra nuisance parameters controlling the mapping between X-ray flux and detection significance, and hence the X-ray selection function from the direct fit to the data (red), the sampling of that fit with the 2D number counts (blue), with the X-ray number counts (green), and the optical number counts (orange). Shifts of the contours with respect to the constraints from the data alone are indicative of residual systematics.

equation (19, red) or adding different number counts likelihoods (2D in blue, X-ray in green, and optical in orange). In principle, we expect no extra information from the number counts on the scaling governing the X-ray selection function. Yet the posterior of the X-ray number counts in particular display shifts compared to the direct fit. This might hint at unresolved systematic effects in the X-ray number counts. Indeed, we find that the X-ray number counts predict a smaller intrinsic scatter σ_X and a steeper mass slope than both the SPT-SZ cross-calibration and the 2D number counts. While at the current stage these putative systematics are smaller than the statistical uncertainties, the empirical methods here already prove to be potent tools for validating the number counts. We plan to include such tests as unblinding conditions for the forthcoming cosmological analysis of this catalogue.

6.2 Outcome of the validation

As outlined in Section 5.3, different methods with different sensitivities to the selection function provide statistically consistent masses. This provides strong evidence for the adequacy of the selection functions we construct in this work. Interestingly, however, non-significant tensions appear on different parameters, mainly in the scaling relation parameters derived from 1D X-ray number counts and 2D optical and X-ray number counts. These tensions are also visible in the comparison of the predicted and the measured number counts (Section 5.2), as well as in the comparison of the inferred masses (Section 5.3). We identify two main scenarios: low intrinsic scatter and steep luminosity–mass trend – preferred by X-ray number counts, and large intrinsic scatter and shallow slope – preferred by 2D number counts. In the following, we will discuss evidence for these two scenarios.

Comparison to the literature does not provide clear guidance on which scenario is more plausible, as can be seen in Fig. 17. The low scatter scenario is in very good agreement with the results

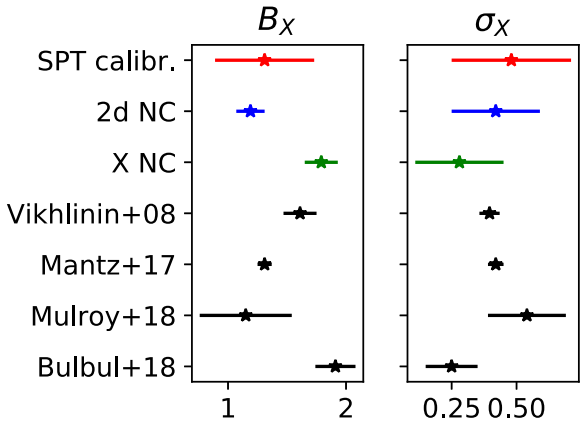

Figure 17. Compilation of results on the mass trend B_X and the intrinsic scatter σ_X of the luminosity–mass relation, compared to our results. While all our results lay within the dispersion of the literature results, this dispersion among the results is larger than the reported error bars, indicating that no consensus has yet been reached.

from Bu19 on *XMM* luminosities of SPT-SZ-selected clusters. On the other hand, weak-lensing calibrated measurements of the luminosity–mass relation on RASS-selected clusters by Mantz et al. (2015) and Mulroy et al. (2019) find shallower mass trends and larger intrinsic scatter in good agreement with our large scatter scenario. In analysing number counts of RASS-selected clusters with X-ray mass information, Vikhlinin et al. (2009) found a mass trend and scatter value consistent with both scenarios.

Further evidence for the amount of intrinsic scatter can be obtained by comparing different measurements of the luminosities. K19 show that there is significant scatter among the luminosities measured by Boller et al. (2016) and those reported by Piffaretti et al. (2011). Namely, a lognormal scatter of 0.48 ± 0.05 for $0.15 < z < 0.3$ and 0.40 ± 0.10 for $0.3 < z$. This in unsettling, considering that the luminosities reported by Piffaretti et al. (2011) are measured on the same *ROSAT* data as the ones by Boller et al. (2016). Given that this effect might be partially sourced by the fixed aperture measurements by Boller et al. (2016), we cannot exclude that the X-ray flux measurement introduces mass-dependent trends. Further investigation of the systematics in the flux measurement methods is clearly required.

The hypothesis of larger scatter in the X-ray mass scaling is further supported by the constraints on the SPT-SZ incompleteness derived from the different posteriors (see Section 5.4.1). Compared to the literature priors, which prefer small scatter but predict high incompleteness, both the SPT-SZ cross-calibration and the 2D number counts predict incompletenesses consistent with zero, mainly due to the larger X-ray intrinsic scatter. On the other hand, the mass calibrations of the SZE–mass scaling determined using different, independent methods (Chiu et al. 2018; Capasso et al. 2019a; Dietrich et al. 2019; Stern et al. 2019) match with the masses emerging from a fully self-consistent cosmological analysis of the SPT-SZ cluster sample (Bocquet et al. 2015; de Haan et al. 2016; Bocquet et al. 2019). In the presence of high incompleteness, this agreement would be coincidental. Larger X-ray scatter is thus made even more plausible, because it predicts low SPT-SZ incompleteness.

In summary, the large scatter/shallower mass trend scenario is supported by the comparison of different luminosity measures, different literature results and the implications of these scenarios on the inferred SPT-SZ incompleteness. Furthermore, we find that the 2D number count fits introduce less internal tension on the parameters of the significance-flux scaling governing the X-ray selection function.

6.3 Impact of the optical incompleteness

As shown throughout this work, we model the selection of the MARD-Y3 sample in a two staged approach, which mirrors the operational creation of the catalogue: (1) we determine an X-ray selection function based on the fact that the candidate catalogue is selected with a X-ray detection significance threshold, and (2) we model the optical cleaning, which is operationally equivalent to a redshift-dependent minimum value for the measured richness. The two result in selection functions in the space of X-ray flux and richness, respectively (cf. Section 2.3).

For ease of representation, we utilize the observable–mass scaling relation to transform these observable selection functions into mass selection functions. This introduces systematic uncertainty through the widths of the posteriors on the scaling relation parameters. The mass selection functions in three redshift bins are shown in Fig. 18. As stated above, the X-ray selection is dominant at most masses. Yet, the optical cleaning introduces an excess incompleteness at the lowest masses, leading to a suppression of the selection probability at those masses.

The fact that the optical selection cannot be completely ignored can be appreciated also from Figs 11 and 12. Given that in these plots we show the number of clusters also as a function of measured richness, we can appreciate that the MARD-Y3 sample displays a sharp, redshift-dependent cut in measured richness. This is the result of the optical cleaning process, which takes effect before the X-ray selection probability nears zero.

The fact that we can consistently infer the masses when marginalizing over a fiducial cosmology indicates that the two stage selection function modelling is adequately describing the sample. This in turn means that optical cleaning with MCMF can provide clean cluster samples even from highly contaminated candidate samples. At the cost of tracking an extra scaling relation, the richness–mass relation, this has the potential to significantly lowering the limiting mass of ongoing and future surveys with SPT, eROSITA, or similar ICM observable-based surveys while maintaining a similar contamination level. Given that all selected clusters in such samples would have a richness in addition to an X-ray or SZE observable, the richness–mass relation would be calibrated along side the X-ray or SZE observable in the context of a direct mass calibration as we have demonstrated with our SPT-SZ cross-calibration. Furthermore, the possibility to perform number counts not only in the X-ray or SZE observable, but in richness alone, or even in the combination of multiple observables, provides additional consistency checks that could be used to reveal unappreciated systematics.

6.4 Implications for cosmological studies

In this work, we explore several techniques that allow us to validate the selection function of a cluster survey. However, we would like to caution that in this work we never directly determine the masses of our clusters. This would require either the measurement of the weak-lensing signal around our clusters, or the study of the projected phase space distribution of spectroscopically observed cluster members. From a formal perspective, such studies can be treated analogously to our SPT-SZ cross-calibration. They will allow us to determine the parameters of the scaling relations to high accuracy, enabling the use of the number counts to study cosmology.

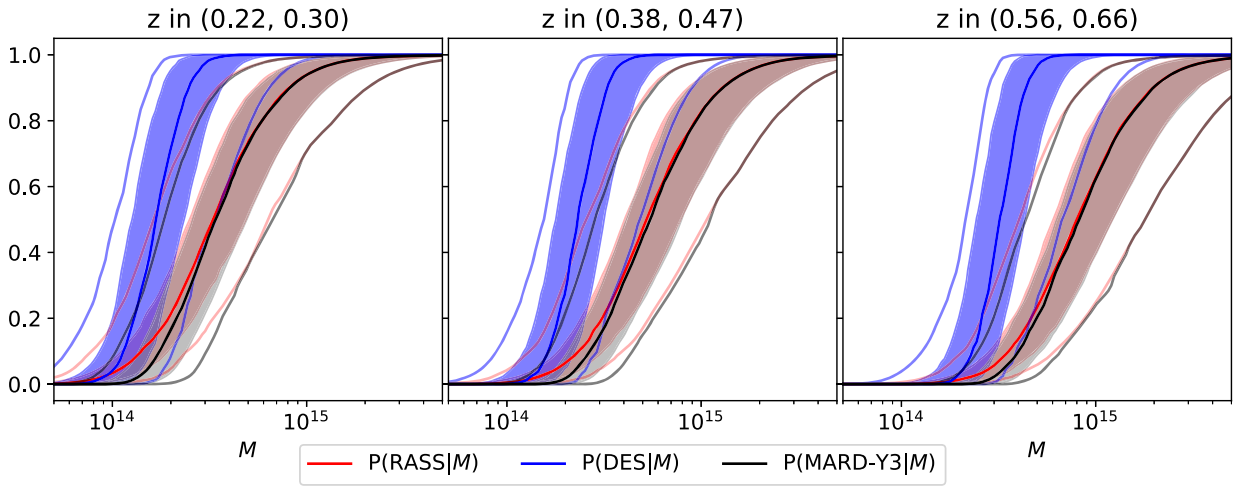

Figure 18. X-ray (red), optical (blue), and combined (black and grey) selection functions as functions of mass for different redshift bins, plotted with the systematic uncertainties derived from the 2D number counts posterior on the scaling relation parameters. The full lines are the median values, the filled region covers the range from the 16th to 84th percentile and the transparent lines show the 2.5th and 97.5th percentile. All masses refer to spherical over densities 500 times the critical density of the Universe. While the combined selection of the sample is clearly dominated by the X-ray selection function at most masses, the optical cleaning introduces some extra incompleteness at low masses, especially at low redshift.

In contrast, our current work assumes the cosmology derived by Bo19 in order to determine the scaling relation parameters from the number counts of the MARD-Y3 sample. Also, the indirect mass information we use in form of the priors on the SZE–mass relation were derived by Bo19 in the same analysis. So they, too, are contingent upon that analysis. The consistency of their result with our modelling is supported by the fact that we do not find a significant level of SPT-SZ incompleteness.

Our work then demonstrates several techniques that we anticipate will be important for controlling systematics in future X-ray-selected cluster samples, especially the sample detected by eROSITA (Predehl et al. 2010; Merloni et al. 2012). First, we have shown that the X-ray selection function can be determined empirically from the selected sample (see Fig. 1). As such, the simplistic assumptions made in forecast works (e.g. Grandis et al. 2019) can easily be replaced by a more accurate description without introducing much numerical complexity. The empirical determination of the selection function also allows one to check for unresolved systematic effects, as demonstrated in Section 2.3.2. As an addition to the set of systematics tests, such techniques are likely to improve the systematics control within eROSITA cluster cosmological studies.

Our work also highlights the use of secondary mass proxies to inform the number counts experiment. We demonstrate that performing the number counts in optical richness despite the X-ray selection provides a valuable source of mass information. In the presence of a direct mass calibration, that mass information would be provided externally, and optical number counts would provide independent cosmological constraints. This in turn allows one to set up another important consistency check, ensuring a higher level of systematics control. On the same note, we also clearly demonstrate the value of additional mass proxies to put direct constraints on the scatter. Indeed, the analysis of the number counts in X-ray flux and richness space was central to revealing the larger scatter in X-ray observable. Given the planned application of MCMF to eROSITA such multi-observable number counts experiments can be undertaken also in that context.

Furthermore, we present here an expansion of earlier work by S15 on detection probabilities of clusters selected by one survey in another survey. Our formalism tests the selection functions of different surveys against each other and thereby gains precious empirical constraints on those selection functions. This method depends on the shape of the mass function for the Eddington bias correction, and on the redshift–distance relation for the X-ray scaling relation. Importantly, however, it is independent of the distribution of clusters in observable and redshift. In turn, these are the major sources of cosmological information in the number counts experiment. Consequently, in the presence of direct mass information to constrain the scaling relation parameters, this technique provides a selection function test that is insensitive to the predicted number of clusters and its redshift evolution. As such this test is ideally suited to validated cluster number count experiments.

Our approach would not only benefit the systematics control in future X-ray and SZE surveys, but also future optical surveys. The selection function in optical surveys remains a source of systematic uncertainty that has been mainly studied through simulations (Costanzi et al. 2019). Applying techniques like ours to empirically validate an optical survey cluster selection function offers important advantages and will become more relevant with the upcoming next generation surveys from Euclid and LSST.

7 CONCLUSIONS

We perform a multiwavelength analysis of the MARD-Y3 sample (K19). This sample was selected by performing an optical follow-up of the X-ray selected 2RXS catalogue (Boller et al. 2016) using DES-Y3 data. The optical followup was carried out using MCMF (Klein et al. 2018), which is a tool that includes spatial and colour filters designed to identify optical counterparts of ICM selected cluster candidates and to exclude random superpositions of X-ray and optical systems. The multiwavelength data set enables for an extensive set of cross-checks and systematics probes of the MARD-Y3 sample, its selection function and the associated observable–mass scaling relations.

We model the selection function (see Section 2.3) of the MARD-Y3 sample as the combination of the X-ray selection function of the candidate sample together with a model of the incompleteness

introduced by the optical cleaning of that sample. We then proceed to calibrate the X-ray luminosity–mass and optical richness–mass relation using different sources of mass information to test whether there is tension in the data set or a flaw in the selection function.

First, we cross-match the MARD-Y3 and the SZE-selected SPT-SZ cluster samples, and calibrate the MARD-Y3 scaling relations using the published calibration of the SZE signal-to-noise–mass relation (see Section 5.1). Second, assuming priors on the cosmological parameters from the most recent SPT-SZ cluster cosmology analysis (Bocquet et al. 2019), we calibrate the observable mass scaling relations from the number counts of MARD-Y3 clusters (see Section 5.2). In addition to the traditional number counts as a function of X-ray flux and redshift, we also use the number counts as a function of richness and redshift and the number counts as a function of X-ray flux, richness, and redshift.

We find that the different flavours of number counts provide scaling relation constraints that are statistically consistent with the constraints from the SPT-SZ calibration performed on the cross-matched sample. This validates the MARD-Y3 selection function, because the SPT-SZ calibration is independent of the MARD-Y3 selection function, while the number count experiments are highly sensitive to it. This leads us to the main conclusion of this work: *optical cleaning with MCMF allows* one to create a clean cluster sample with a controllable selection function. Once direct mass information is available, we will be able to study cosmology using the MARD-Y3 number counts. The fact that the incompleteness (primarily at low masses) introduced by optical cleaning can be modelled using the richness–mass relation implies that much larger, reliable cluster samples extending to higher redshift and lower masses can be constructed from ICM-based surveys if appropriately deep optical and NIR data are available.

In these tests we identify some moderate tension between constraints on the luminosity–mass relation from X-ray number counts and 2D (optical+X-ray) number counts: while the former prefers small intrinsic X-ray scatter and a steep mass trend, the latter prefers a shallower mass trend and larger intrinsic scatter. This hints at some unresolved systematic on the X-ray side. As discussed in Section 6.2, the high scatter scenario is supported by the scatter among different measurements of luminosity on the same X-ray raw data highlighted in K19, a further indicator of systematics in the flux measurement. Nevertheless, the individual masses derived from the different scenarios are consistent within the uncertainties. Because there is no consensus in the literature, this question merits further investigation once direct mass information is available.

In Section 5.3, we present the implications for MARD-Y3 masses from different scaling relations that emerge from the tests described above. There is a tendency for these masses to lie below those calculated using externally calibrated relations from the literature (Saro et al. 2015; Bulbul et al. 2019), and the largest tensions occur at low masses.

We also study the MARD-Y3 selection function by comparing the matched and unmatched MARD-Y3 clusters in the SPT-SZ sample and vice versa. If the selection functions for MARD-Y3 and SPT-SZ are well understood then the number of matched and unmatched clusters should be fully consistent with the statistical expectations. Simply stated, this test allows us to constrain MARD-Y3 contamination or SPT-SZ incompleteness (the two effects are degenerate in this test). As discussed in Section 5.4.1, in the large scatter luminosity–mass scenario, we find no evidence for either effect, while in the low scatter scenario we find evidence at the 2σ level for either contamination or incompleteness. Given that the MARD-Y3 sample contamination is estimated to be 2.5 per cent (K19) and given that the SPT-SZ sample has been used to produce cosmological constraints in good agreement with independent probes (de Haan et al. 2016; Bocquet et al. 2019), we take this as further evidence supporting the large scatter scenario.

Looking at the probability of an MARD-Y3 confirmation of an SPT-SZ-selected cluster we find a subsample of clusters whose SZE properties suggest they should not have been detected in MARD-Y3, but they are. The size of this sample is susceptible to the scaling relation constraints assumed. As discussed in Section 5.4.2, if we model this as an outlier fraction in the distribution of scatter about the mass–observable relations (either abnormally high X-ray flux and richness, or low SZE signature), we find a preference for an outlier fraction of $\sim$5 per cent–10 per cent with a detection significance ranging from 1.4σ to 3.6σ, depending on the scaling relation constraints assumed. More accurate and independent mass information is needed to further elucidate this aspect of the cluster population.

From a methodological perspective we demonstrate several new techniques:

(i) Optical follow-up enables for three different flavours of number counts. While we demonstrate the potential of multi-observable number counts, the real novelty is that one can perform number counts as a function of optical richness for a predominantly X-ray-selected sample in a consistent manner. In a blinded WL-calibrated cosmological analysis we would demand that the blinded cosmology from these three likelihoods be consistent.

(ii) We improve the technique of studying matched and unmatched clusters in two independent samples by including binomial statistics and marginalizing over the systematic uncertainties associated with lack of knowledge of the observable–mass relation parameters. With the use of probability trees, extra probabilities, such as those quantifying contamination, incompleteness or outlier fractions, can all be constrained in a statistically sound way. This technique does not depend on the amplitude and redshift evolution of the number of objects, reducing its cosmological sensitivity.

(iii) We present a flexible empirical method to determine the X-ray selection function from the data itself (see Fig. 1). It does not require any assumptions about cluster morphology. The empirical nature of the constraint also marginalizes over the inherent uncertainty of the selection function by sampling extra nuisance parameters. Shifts in these nuisance parameters when, for example, calibrating the observable–mass relation using different sources of information can serve as a further test of systematic.

The techniques highlighted here have the potential to enable better control of systematic effects in cosmological studies of current and upcoming cluster surveys. They also demonstrate the potential of multiwavelength analysis of cluster samples not only to inform the selection function modelling of individual surveys, but also to identify interesting cluster populations. This will help exploit the wealth of information provided by deep and wide surveys in X-ray, optical, NIR, and millimetre wavelengths.

ACKNOWLEDGEMENTS

We acknowledge financial support from the MPG Faculty Fellowship program, the DFG Cluster of Excellence ‘Origin and Structure of the Universe’, the new DFG cluster ‘Origins’, and the Ludwig-Maximilians-Universität Mnchen. Numerical computations in this work relied on the PYTHON packages NUMPY (Travis 2006) and SCIPY (Jones et al. 2001). The plots were produced using the package MATPLOTIB (Hunter 2007). The marginal contour plots were created using PYGTC (Bocquet & Carter 2016). Posterior samples have been

drawn from the likelihood functions and the priors using EMCEE (Foreman-Mackey et al. 2013).

This work was performed in the context of the South-Pole Telescope scientific program. SPT is supported by the National Science Foundation through grant PLR-1248097. Partial support is also provided by the NSF Physics Frontier Center grant PHY-0114422 to the Kavli Institute of Cosmological Physics at the University of Chicago, the Kavli Foundation, and the Gordon and Betty Moore Foundation grant GBMF 947 to the University of Chicago. This work is also supported by the U.S. Department of Energy. AAS acknowledges support by US NSF AST-1814719.

This paper has gone through internal review by the DES collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the DES.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union 7th Framework Programme (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under contract no. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

DATA AVAILABILITY

There are no new data associated with this article. See Section 4.1 for references to the cluster catalogues we used.

REFERENCES

Albrecht A. et al., 2006, preprint (arXiv:astro-ph/0609591)
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Andersson K. et al., 2011, ApJ, 738, 48
Applegate D. E. et al., 2014, MNRAS, 439, 48
Bardeau S., Soucail G., Kneib J.-P., Czoske O., Ebeling H., Hudelot P., Smail I., Smith G. P., 2007, A&A, 470, 449
Bleem L. E. et al., 2015, ApJS, 216, 27
Bocquet S., Carter F. W., 2016, J. Open Source Softw., 1, 46
Bocquet S. et al., 2015, ApJ, 799, 214
Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
Bocquet S. et al., 2019, ApJ, 878, 55
Böhringer H. et al., 2001, A&A, 369, 826
Boller T., Freyberg M. J., Trümper J., Haberl F., Voges W., Nandra K., 2016, A&A, 588, A103
Bulbul E. et al., 2019, ApJ, 871, 50
Capasso R. et al., 2019a, MNRAS, 482, 1043
Capasso R. et al., 2019b, MNRAS, 486, 1594
Chiu I. et al., 2018, MNRAS, 478, 3072
Clerc N. et al., 2014, MNRAS, 444, 2723
Clerc N. et al., 2018, A&A, 617, A92
Costanzi M. et al., 2019, MNRAS, 482, 490
Dark Energy Survey Collaboration, 2016, MNRAS, 460, 1270
de Haan T. et al., 2016, ApJ, 832, 95
DES Collaboration, 2018, ApJS, 239, 18
Dietrich J. P. et al., 2019, MNRAS, 483, 2871
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Grandis S., Mohr J. J., Dietrich J. P., Bocquet S., Saro A., Klein M., Paulus M., Capasso R., 2019, MNRAS, 488, 2041
Haiman Z., Mohr J. J., Holder G. P., 2001, ApJ, 553, 545
Hasselfield M. et al., 2013, J. Cosmol. Astropart. Phys., 7, 008
Hennig C. et al., 2017, MNRAS, 467, 4015
Hilton M. et al., 2018, ApJS, 235, 20
Hincks A. D. et al., 2010, ApJS, 191, 423
Hoekstra H., Mahdavi A., Babul A., Bildfell C., 2012, MNRAS, 427, 1298
Hu W., 2003, Phys. Rev. D, 67, 081304
Huang N. et al., 2020, AJ, 159, 110
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Israel H., Reiprich T. H., Erben T., Massey R. J., Sarazin C. L., Schneider P., Vikhlinin A., 2014, A&A, 564, A129
Ivezic Z. et al., 2019, ApJ, 873, 111
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jones E. et al., 2001, SciPy: Open Source Scientific tools for Python. Available at: http://www.scipy.org/, (December 2019)
Klein M. et al., 2018, MNRAS, 474, 3324
Klein M. et al., 2019, MNRAS, 488, 739
Koester B. P. et al., 2007, ApJ, 660, 239
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Liu J. et al., 2015, MNRAS, 448, 2085
Majumdar S., Mohr J. J., 2003, ApJ, 585, 603
Majumdar S., Mohr J. J., 2004, ApJ, 613, 41
Mantz A., Allen S. W., Rapetti D., Ebeling H., 2010, MNRAS, 406, 1759
Mantz A. B. et al., 2015, MNRAS, 446, 2205
Mantz A. B. et al., 2016, MNRAS, 463, 3582
McClintock T. et al., 2019a, MNRAS, 482, 1352

McClintock T. et al., 2019b, ApJ, 872, 53
McDonald M. et al., 2014, ApJ, 794, 67
Melchior P. et al., 2015, MNRAS, 449, 2219
Melchior P. et al., 2017, MNRAS, 469, 4899
Merloni A. et al., 2012, preprint (arXiv:1209.3114)
Mulroy S. L. et al., 2019, MNRAS, 484, 60
Okabe N., Smith G. P., 2016, MNRAS, 461, 3794
Okabe N., Zhang Y.-Y., Finoguenov A., Takada M., Smith G. P., Umetsu K., Futamase T., 2010, ApJ, 721, 875
Pacaud F. et al., 2006, MNRAS, 372, 578
Piffaretti R., Arnaud M., Pratt G. W., Pointecouteau E., Melin J.-B., 2011, A&A, 534, A109
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration XXVII, 2016, A&A, 594, A27
Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., 2019, Space Sci. Rev., 215, 25
Predehl P. et al., 2010, in Arnaud M., Murray S. S., Takahashi T., eds, Proc. SPIE Vol. 7732, Space Telescopes and Instrumentation 2010: Ultraviolet to Gamma Ray. SPIE, Bellingham, p. 77320U
Rigault M. et al., 2018, preprint (arXiv:1806.03849)
Romer A. K., Viana P. T. P., Liddle A. R., Mann R. G., 2001, ApJ, 547, 594
Rykoff E. S. et al., 2016, ApJS, 224, 1
Saro A. et al., 2015, MNRAS, 454, 2305
Schrabback T. et al., 2018, MNRAS, 474, 2635
Sifón C. et al., 2013, ApJ, 772, 25
Stern C. et al., 2019, MNRAS, 485, 69
Sunyaev R. A., Zeldovich Y. B., 1972, Comments Astrophys. Space Phys., 4, 173
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Travis E., 2006, A Guide to NumPy. Trelgol Publishing, USA
Vanderlinde K. et al., 2010, ApJ, 722, 1180
Vikhlinin A., McNamara B. R., Forman W., Jones C., Quintana H., Hornstrup A., 1998, ApJ, 502, 558
Vikhlinin A. et al., 2009, ApJ, 692, 1033
Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, ApJ, 646, 881
Werner N., Churazov E., Finoguenov A., Markevitch M., Burenin R., Kaastra J. S., Böhringer H., 2007, A&A, 474, 707
Zhang Y.-Y. et al., 2017, A&A, 599, A138

APPENDIX A: X-RAY FLUX ERROR MODEL

As outlined in Section 2.2 in some application it is not sufficient to know the measurement uncertainty only for the objects in the catalogue, but the measurement uncertainty is also needed for arbitrary values of measured flux $\hat{f}_X$ and redshift z. We therefore seek to predict $\hat{\sigma}^2_X(\hat{f}_X, z, t_{exp})$ from the measured entries $\hat{\sigma}^{(i)}_X$. First we note that the measurement uncertainties in the catalogue scale with the exposure time approximately like $\hat{\sigma}^{(i)}_X \sim t^{-0.5}_{exp}$. We thus bin the quantity $(\hat{\sigma}^{(i)}_X)^2\, t_{exp}/400s$ in fine redshift and measured flux bins, as shown in the upper panel of Fig. A1. This is then extrapolated and

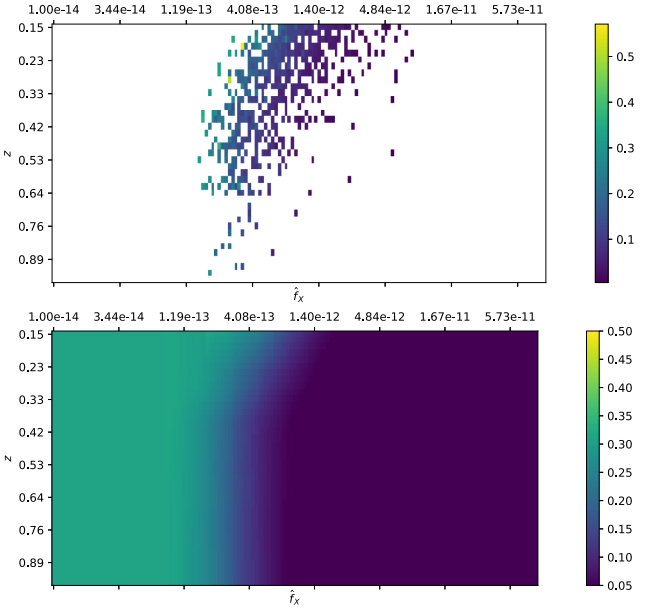

Figure A1. Construction of the expected measurement uncertainty at the median redshift (lower panel) as a function of redshift (y-axis) and measured flux (x-axis), from the normalized measurement uncertainties reported in the catalogue (upper panel). In the range where we have data, the predicted observational measurement uncertainty nicely extrapolates the trends in the catalogue.

smoothed to provide a prediction of the measurement uncertainty $\sigma^2_{pred}(\hat{f}_X, z)$ at each measured flux $\hat{f}_X$ and redshift z, if the exposure time was $t_{exp} = 400\,s$, shown in the lower panel of Fig. A1. This prediction can than be scaled to the desired exposure time assuming the scaling above, that is

$$\hat{\sigma}^2_X(\hat{f}_X, z, t_{exp}) = \sigma^2_{pred}(\hat{f}_X, z)\frac{400s}{t_{exp}}. \quad \text{(A1)}$$

Applying this prediction the cluster in our catalogue and comparing the resulting uncertainties to the actual measurement uncertainty leads to a mean relative error of 5.6 per cent. Furthermore, these residuals display no strong trends with background brightness, neutral hydrogen column density or measured extent. Given the small magnitude, we choose to ignore this source of systematic uncertainty, which could be included at the cost of sampling extra nuisance parameters.

APPENDIX B: GALLERY OF MULTIWAVELENGTH CLUSTER IMAGES

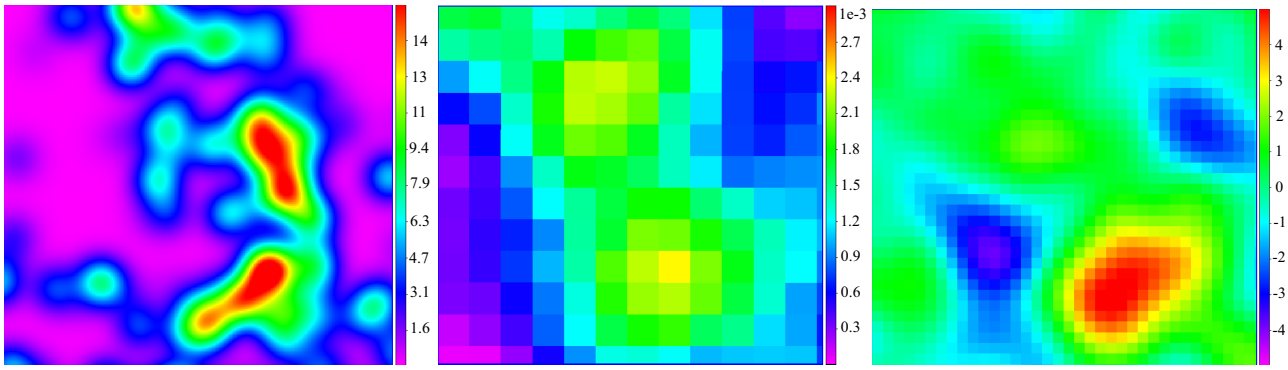

Figure B1. SPT-CL J2358-6129. Red sequence galaxy density map in DES at $z = 0.403$ (left-hand panel), exposure and background corrected count rate from RASS (centre panel), and SPT signal-to-noise map in the smaller filter scale (right-hand panel). One pixel in the RASS count rate maps has diameter 45 arcsec, all images have the same size. The two MARD-Y3 detected clusters (X-ray peaks in the centre panel) are matched to the same SPT-SZ detection (right-hand panel, cf. Section 4.1.3).

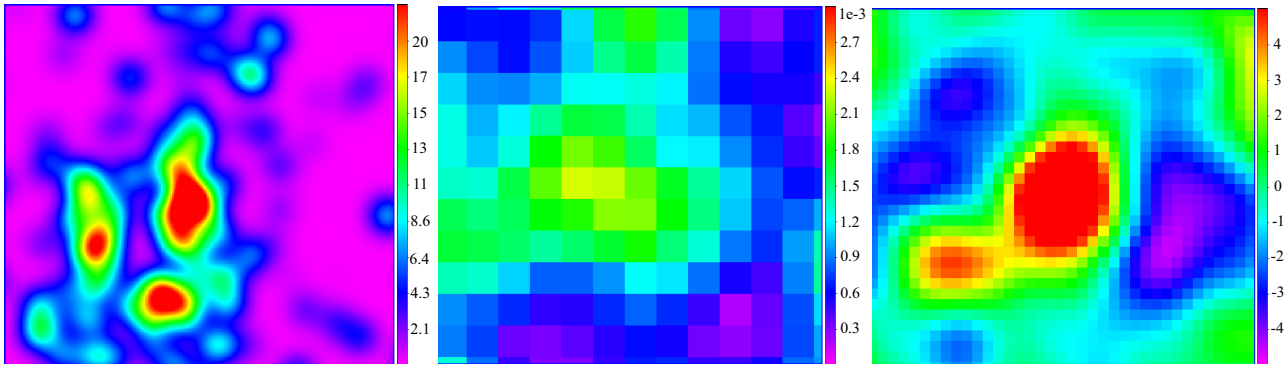

Figure B2. SPT-CL J2331-5051 and SPT-CL J2331-5053. Red sequence galaxy density map in DES at $z = 0.577$ (left-hand panel), exposure and background corrected count rate from RASS (centre panel), and SPT signal-to-noise map in the smaller filter scale (right-hand panel). One pixel in the RASS count rate maps has diameter 45 arcsec, all images have the same size. The two SPT-SZ detected clusters (right-hand panel) are matched to the same MARD-Y3 detection (right-hand panel, cf. Section 4.1.3).

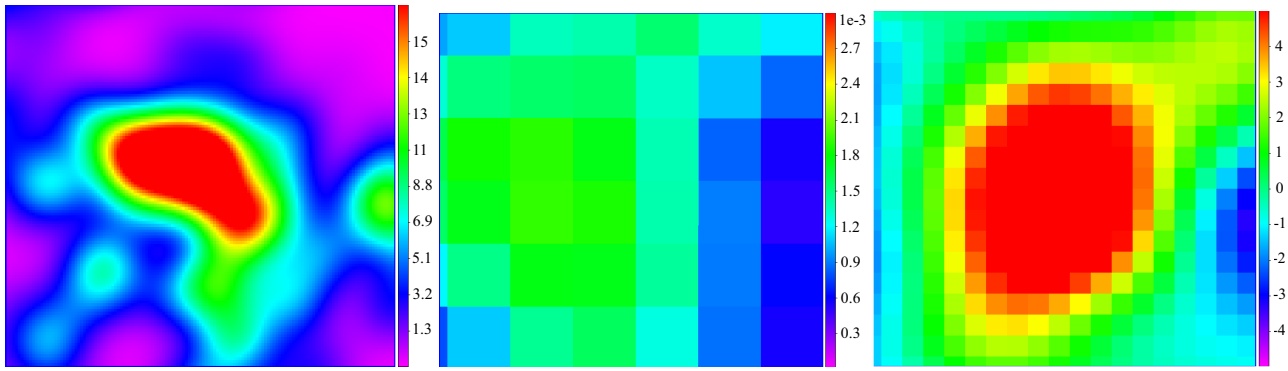

Figure B3. SPT-CL J0218-4233. Red sequence galaxy density map in DES at $z = 0.755$ (left-hand panel), exposure and background corrected count rate from RASS (centre panel), and SPT signal-to-noise map in the smaller filter scale (right-hand panel). One pixel in the RASS count rate maps has diameter 45 arcsec, all images have the same size. This SPT-SZ detection was unexpectedly confirmed by MARD-Y3, cf. Section 5.4.2.

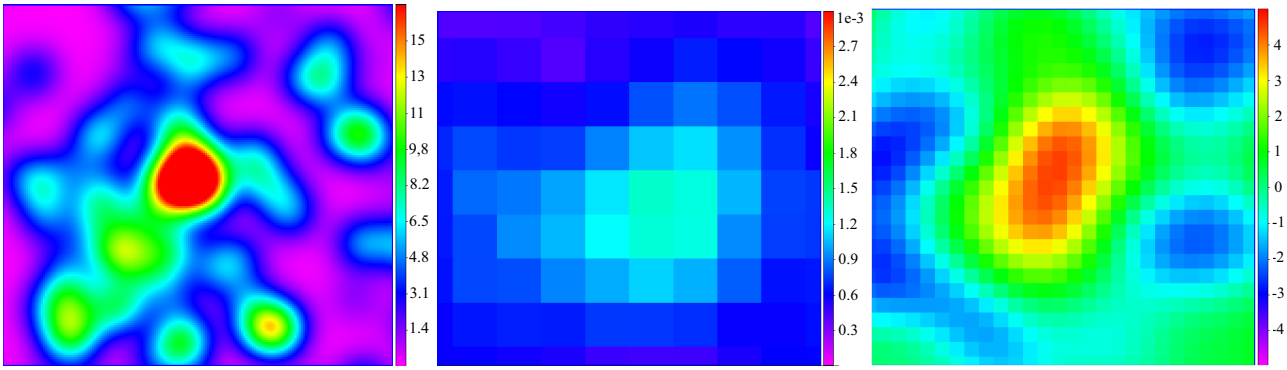

Figure B4. SPT-CL J0324-6236. Red sequence galaxy density map in DES at $z = 0.638$ (left-hand panel), exposure and background corrected count rate from RASS (centre panel), and SPT signal-to-noise map in the smaller filter scale (right-hand panel). One pixel in the RASS count rate maps has diameter 45 arcsec, all images have the same size. This SPT-SZ detection was unexpectedly confirmed by MARD-Y3, cf. Section 5.4.2.

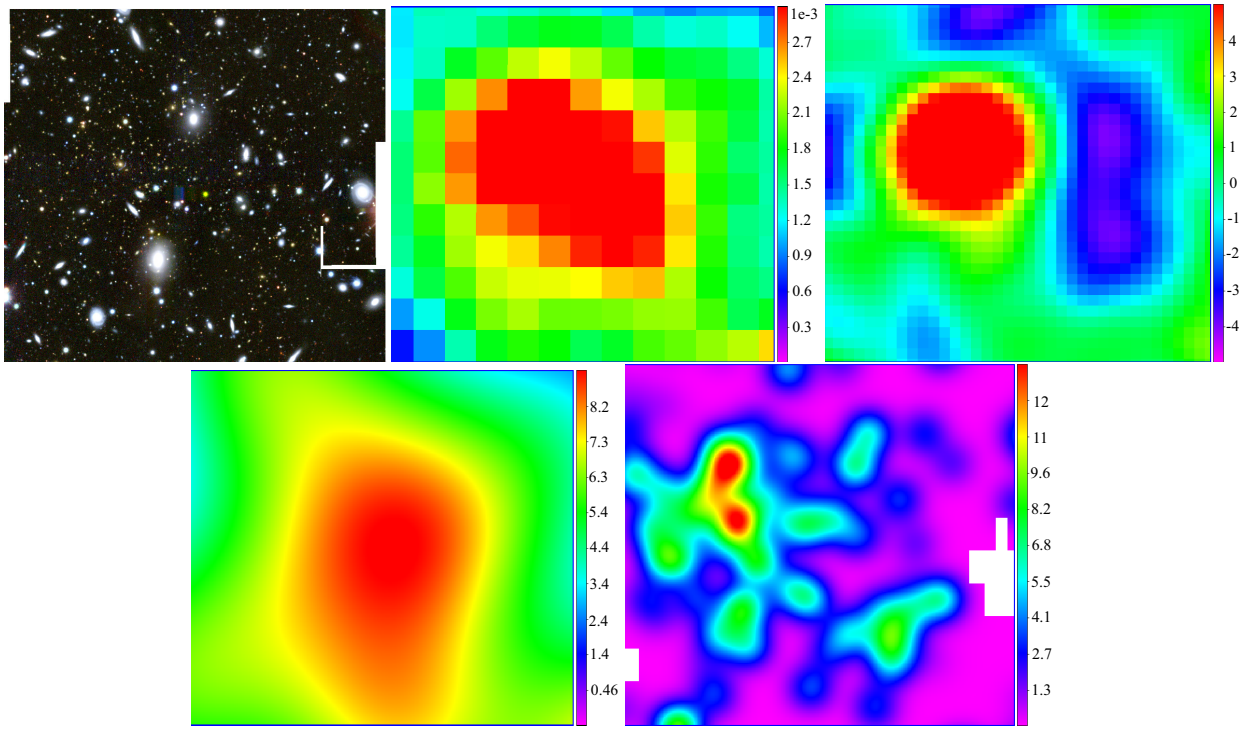

Figure B5. 2RXS J033045.2-522845. DES image (upper left panel), exposure and background corrected count rate from RASS (upper centre panel), and SPT signal-to-noise map in the smaller filter scale (upper right panel), Red sequence galaxy density map in DES at $z = 0.056$ (lower left panel) and at $z = 0.428$ (lower right panel). One pixel in the RASS count rate maps has diameter 45arcsec, all images have the same size. This MARD-Y3 detection was unexpectedly missed by SPT-SZ, cf. Section 5.4.1. This is due to a catastrophic redshift failure of MCMF when run on SPT-SZ detections: it selected the low-redshift group (white galaxies in the upper left panel, red sequence galaxy density in the lower left panel), while the actual structure is at intermediate redshift (galaxy density in the lower right corner). When zooming into DES image, a blue Einstein arc around the brightest central galaxy can be seen.

[1]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[2]*Excellence Cluster Origins, Boltzmannstr. 2, D-85748 Garching, Germany*
[3]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. D-85748, Garching, Germany*
[4]*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*
[5]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP-05314-970, Brazil*
[6]*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ – 20921-400, Brazil*
[7]*Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA*
[8]*Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA*
[9]*Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA*
[10]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[11]*UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, Sorbonne Universités, F-75014 Paris, France*
[12]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK*

[13]*Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, UK*
[14]*Kavli Institute for Particle Astrophysics and Cosmology, P.O. Box 2450, Stanford University, Stanford, CA 94305, USA*
[15]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[16]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Av. Complutense, 40, 28040 Madrid, Spain*
[17]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[18]*National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*
[19]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[20]*Oskar Klein Centre, Department of Physics, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden*
[21]*INAF – Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[22]*IFPU – Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[23]*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ – 20921-400, Brazil*
[24]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ – 20921-400, Brazil*
[25]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[26]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[27]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[28]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[29]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[30]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[31]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[32]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[33]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[34]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[35]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[36]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[37]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[38]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[39]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[40]*Center for Astrophysics, Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[41]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[42]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[43]*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
[44]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[45]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[46]*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
[47]*Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA*
[48]*School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK*
[49]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[50]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK*
[51]*Fakultät für Physik, Ludwig-Maximilians Universität München, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 München, Germany*

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