

# HSC-XXL : Baryon budget of the 136 XXL Groups and Clusters \*

**Daichi AKINO<sup>1</sup>, Dominique ECKERT<sup>2</sup>, Nobuhiro OKABE<sup>1,3,4</sup>, Mauro SERENO<sup>5,6</sup>, Keiichi UMETSU<sup>7</sup>, Masamune OGURI<sup>8,9,10</sup>, Fabio GASTALDELLO<sup>11</sup>, I-Non CHIU<sup>7</sup>, Stefano ETTORI<sup>5,6</sup>, August E. EVRARD<sup>12,13</sup>, Arya FARAH<sup>12,13</sup>, Ben MAUGHAN<sup>14</sup>, Marguerite PIERRE<sup>15</sup>, Marina RICCI<sup>16</sup>, Ivan VALTCHANOV<sup>17</sup>, Ian MCCARTHY<sup>18</sup>, Sean MCGEE<sup>19</sup>, Satoshi MIYAZAKI<sup>20</sup>, Atsushi J. NISHIZAWA<sup>21</sup> and Masayuki TANAKA<sup>20</sup>**

<sup>1</sup>Physics Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan

<sup>2</sup>Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland

<sup>3</sup>Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan

<sup>4</sup>Core Research for Energetic Universe, Hiroshima University, 1-3-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan

<sup>5</sup>INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, I-40129 Bologna, Italy

<sup>6</sup>INFN, Sezione di Bologna, viale Berti Pichat 6/2, 40127 Bologna, Italy

<sup>7</sup>Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

<sup>8</sup>Research Center for the Early Universe, University of Tokyo, Tokyo 113-0033, Japan

<sup>9</sup>Department of Physics, University of Tokyo, Tokyo 113-0033, Japan

<sup>10</sup>Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, Chiba 277-8582, Japan

<sup>11</sup>INAF - IASF Milano, via Bassini 15, I-20133 Milano, Italy

<sup>12</sup>Department of Physics and Michigan Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>13</sup>McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<sup>14</sup>H. H. Wills Physics Laboratory, University of Bristol, Tyndall Ave., Bristol BS8 1TL, UK

<sup>15</sup>AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France

<sup>16</sup>Laboratoire d'Annecy de Physique des Particules, Université Savoie Mont Blanc, CNRS/IN2P3, 74941 Annecy, France

<sup>17</sup>Telespazio UK for ESA, European Space Astronomy Centre, Operations Department, E-28691 Villanueva de la Cañada, Spain

<sup>18</sup>Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

<sup>19</sup>School of Astronomy and Physics, University of Birmingham, Edgbaston, Birmingham, B15, 2TT, UK

<sup>20</sup>National Astronomical Observatory of Japan, Osawa 2-21-1, Mitaka, Tokyo 181-8588, Japan

<sup>21</sup>Institute for Advanced Research, Nagoya University Furocho, Chikusa-ku, Nagoya,

464-8602 Japan

\*E-mail: m204964@hiroshima-u.ac.jp, Dominique.Eckert@unige.ch, okabe@hiroshima-u.ac.jp

Received ; Accepted

## Abstract

We present our determination of the baryon budget for an X-ray-selected XXL sample of 136 galaxy groups and clusters spanning nearly two orders of magnitude in mass ( $M_{500} \sim 10^{13} - 10^{15} M_{\odot}$ ) and the redshift range  $0 \lesssim z \lesssim 1$ . Our joint analysis is based on the combination of HSC-SSP weak-lensing mass measurements, XXL X-ray gas mass measurements, and HSC and SDSS multiband photometry. We carry out a Bayesian analysis of multivariate mass-scaling relations of gas mass, galaxy stellar mass, stellar mass of brightest cluster galaxies (BCGs), and soft-band X-ray luminosity, by taking into account the intrinsic covariance between cluster properties, selection effect, weak-lensing mass calibration, and observational error covariance matrix. The mass-dependent slope of the gas mass–total mass ( $M_{500}$ ) relation is found to be  $1.29^{+0.16}_{-0.10}$ , which is steeper than the self-similar prediction of unity, whereas the slope of the stellar mass–total mass relation is shallower than unity,  $0.85^{+0.12}_{-0.09}$ . The BCG stellar mass weakly depends on cluster mass with a slope of  $0.49^{+0.11}_{-0.10}$ . The baryon, gas mass, and stellar mass fractions as a function of  $M_{500}$  agree with the results from numerical simulations and previous observations. We successfully constrain the full intrinsic covariance of the baryonic contents. The BCG stellar mass shows the larger intrinsic scatter at a given halo total mass, followed in order by stellar mass and gas mass. We find a significant positive intrinsic correlation coefficient between total (and satellite) stellar mass and BCG stellar mass and no evidence for intrinsic correlation between gas mass and stellar mass. All the baryonic components show no redshift evolution.

**Key words:** Galaxies: clusters: intracluster medium - X-rays: galaxies: clusters - Gravitational lensing: weak - Galaxies: stellar content

## 1 Introduction

Galaxy groups and clusters are self-gravitating objects with total mass between  $\sim 10^{13} M_{\odot}$  and  $\sim 10^{15} M_{\odot}$ . They contain diffuse thin plasma, galaxies and dark matter. The diffuse gas, referred to as hot baryon, is observed by X-ray satellites or Sunyaev-Zel'dovich (SZ) effect. The cold baryons reside mostly in galaxies, whose stellar component can be observed by optical and/or (near)-infrared telescopes. Galaxies are mainly classified as central brightest cluster galaxies (BCGs) or satellite galaxies. Based on the hierarchical structure formation model, objects form via gravitational collapse of a large volume and thus collect baryons into their halo potentials. Therefore, in the absence of dissipation, the baryon mass fraction is expected to be close to the universal average,  $\Omega_b/\Omega_m$ , measured from cosmic microwave background (CMB) experiments (e.g. White et al. 1993; Evrard 1997; Ettori 2003). Moreover, the cold and hot baryons affect each other through non-gravitational interactions such as star formation, cluster mergers, and energy feedback process by active galactic nuclei (AGNs) and supernove

(SN).

Recent numerical simulations (e.g. Young et al. 2011; McCarthy et al. 2011; Planelles et al. 2013; Martizzi et al. 2014; Le Brun et al. 2014; Wu et al. 2015; Sembolini et al. 2016a; McCarthy et al. 2017; Barnes et al. 2017b; Farahi et al. 2018a; Henden et al. 2020; Farahi et al. 2020) showed that the radiative processes convert gas to stars and significantly affect the evolution of the baryonic components. The details highly depend on AGN models and radiative codes (e.g. McCarthy et al. 2011; Le Brun et al. 2014; Sembolini et al. 2016b). In general, the star formation rate is more efficient in group scale of  $\sim 10^{13} M_{\odot}$  than in massive clusters of  $\sim 10^{15} M_{\odot}$ . Furthermore, AGN feedback in groups is energetic enough to expel hot gas out from their relatively shallow gravitational potentials. It is expected that the baryon contents depend on the halo mass. Therefore, a cluster sample covering as wide mass range as possible provides us with a unique opportunity to understand baryonic physics and its relationship with cluster properties.

The XXL Survey (Pierre et al. 2016; Pacaud et al. 2016; Giles et al. 2016; Lieu et al. 2016; Pompei et al. 2016; Adami et al. 2018; Guglielmo et al. 2018) is one of the largest observ-

\* Based on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan.

ing program undertaken by *XMM-Newton*, covering two distinct sky areas for a total of 50 square degrees down to a sensitivity of  $6 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$  for point-like sources ([0.5–2] keV band). Nearly four hundreds galaxy clusters and groups have been detected (Adami et al. 2018) over a wide range of nearly two orders of magnitude in mass ( $10^{13} - 10^{15} M_{\odot}$ ) up to  $z \sim 2$ . The XXL cluster sample is optimal (Adami et al. 2018; Eckert et al. 2016; Umetsu et al. 2020; Sereno et al. 2020; Willis et al. 2021) for studying the baryon budget of groups and clusters.

Eckert et al. (2016) investigated the gas mass fraction for 100 clusters and the stellar mass fraction for 34 clusters from the XXL first cluster catalog (DR1; Pcaud et al. 2016). Each cluster mass was estimated through their X-ray temperature, calibrated with weak-lensing masses for a subset of 38 clusters covered by the CFHTLS Survey (Lieu et al. 2016). They found that the total baryon fraction within  $r_{500}$  falls short of  $\Omega_b/\Omega_m$  by about a factor of two. Here, the subscript 500 denotes that the mean enclosed density is 500 times the critical density of the Universe at the cluster redshift.

The Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018a, 2018b; Miyazaki et al. 2018b; Komiyama et al. 2018; Kawanomoto et al. 2018; Furusawa et al. 2018; Bosch et al. 2018; Huang et al. 2018b; Coupon et al. 2018; Aihara et al. 2019; Nishizawa et al. 2020) is an on-going wide-field optical imaging survey composed of three layers of different depths (Wide, Deep and UltraDeep). The Wide layer is designed to obtain five-band (*grizy*) imaging over  $1400 \text{ deg}^2$ . The survey footprint significantly overlaps with the northern sky of the XXL Survey. The HSC-SSP Survey has excellent imaging quality ( $\sim 0.7 \text{ arcsec}$  seeing in *i*-band) and reaches a depth of  $r \lesssim 26 \text{ ABmag}$ , enabling the measurement of weak-lensing masses and photometry of the XXL clusters in the overlapped footprint.

Umetsu et al. (2020) measured weak-lensing masses for the 136 XXL clusters in the HSC-SSP survey footprint using the HSC-SSP shape catalog (see details in Mandelbaum et al. 2018a, 2018b). They found that the CFHTLS weak-lensing masses,  $M_{500}$ , are on average  $34 \pm 20$  percent higher than the HSC-SSP ones. Sereno et al. (2020) studied the multivariate scaling relations of X-ray luminosity, temperature, gas mass, and hydrostatic mass for 118 XXL clusters. They measured the gas mass within an overdensity radius,  $r_{500}^{\text{Eckert}}$ , which is computed with an iterative procedure using the surface brightness profile and the  $f_g - M_{500}$  relation from Eckert et al. (2016). However, they used the HSC-SSP weak-lensing masses ( $M_{500}^{\text{WL}}$ ; Umetsu et al. 2020) and investigated the scaling relation between the gas mass  $M_g(< r_{500}^{\text{Eckert}})$  and the weak-lensing mass  $M_{500}^{\text{WL}}(< r_{500}^{\text{WL}})$  defined at different radii. Furthermore, Sereno et al. (2020) did not consider the mass of stellar components in the scaling relation analysis.

Therefore, it is vitally important to measure the gas mass

and the stellar mass using the same overdensity radius as the weak-lensing mass and investigate the gas mass, stellar mass, and total baryon mass fractions in a self-consistent manner.

The paper investigates the gas and stellar mass fractions of the 136 XXL clusters, which are consistently measured within the overdensity radii  $r_{500}$  determined by weak-lensing masses (Umetsu et al. 2020). We employ Bayesian forward modeling, following Sereno (2016) and Sereno et al. (2020), to study multivariate scaling relations. The error covariance matrix, including error correlation induced by the same apertures of the weak-lensing masses, is fully propagated into the scaling relation analysis. Our analysis considers both selection effect and weak-lensing mass calibration. Data analysis is described in Sec.2, results are presented in Sec.3, discussed in Sec.4, and summarized in Sec.5. The paper adopts cosmological parameters of  $\Omega_{m,0} = 0.28$ ,  $\Omega_{\Lambda,0} = 0.72$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2 Data Analysis

### 2.1 XXL cluster sample

The parent cluster sample consists of spectroscopically confirmed X-ray-selected systems of class C1 and C2 drawn from the XXL second data release (DR2) catalog (see details; Adami et al. 2018). The C1 class has a high purity rate (the fraction of false detections  $\sim 5\%$ ) with respect to spurious detections or blended point sources (Adami et al. 2018). The C2 class consists of fainter, hence less-well characterized objects and allows up to 50% contamination by misclassified point sources. On average, the C2 clusters have lower masses than the C1 at fixed redshift. The C2 clusters used in this study are those clusters that could be spectroscopically confirmed by using currently available galaxy redshifts (from the literature and the XXL spectroscopic surveys). Hence, the C2 selection function is, strictly speaking, currently undefined. This paper uses the 83 C1 and 53 C2 spectroscopically confirmed clusters (for a total of 136 clusters) found in the region of overlap between the HSC-SSP and XXL surveys ( $25 \text{ deg}^2$ ). This is the same sample definition used in Umetsu et al. (2020). Umetsu et al. (2020) measured weak-lensing masses for the 136 XXL clusters using the HSC-SSP shape catalog (Mandelbaum et al. 2018a, 2018b). Only galaxies satisfying the full-color and full-depth criteria from the HSC galaxy catalogue were used for precise shape measurements and photometric redshift estimations. Background galaxies behind each cluster are securely selected by their photometric redshift probability distribution, following Medezinski et al. (2018). The weighted number density of background source galaxies is  $n_{\text{gal}} \simeq 22.1 \text{ arcmin}^{-2}$ . Masses are estimated from posterior probability distributions obtained assuming a Navarro–Frenk–White (hereafter, NFW) density profile (Navarro et al. 1996). The weak-lensing mass range covers from group scales  $M_{500}^{\text{WL}} \sim 10^{13} M_{\odot}$  to cluster

scales  $\sim 6 \times 10^{14} M_{\odot}$ . Including the upper bound of the weak-lensing mass uncertainties, the sample reaches  $10^{15} M_{\odot}$ . Since the weak-lensing (WL) mass measurement for a low mass cluster of  $\mathcal{O}(10^{13} M_{\odot})$  is noisy (with a median weak-lensing  $S/N$  of 1.1), Umetsu et al. (2020) validated the bias and scatter of the weak-lensing mass measurements as a function of true mass through numerical simulations and found a mild underestimation weakly depending on the halo mass. This study uses  $M_{500}^{\text{WL}}$  and  $r_{500}^{\text{WL}}$  measurements and weak-lensing mass calibration from Umetsu et al. (2020), which is described in Sec 2.4.3. We use the XXL centers as cluster centers (Umetsu et al. 2020).

To summarize, we use the 136 XXL clusters with 83 C1 and 53 C2 clusters. The weak-lensing mass range is  $M_{500}^{\text{WL}} \sim 10^{13} - 6 \times 10^{14} M_{\odot}$ . The redshift range covers an interval from 0.031 to 1.033. The average and median redshifts for the entire, C1, and C2 samples are  $\langle z_c \rangle = 0.38, 0.34, 0.45$  and  $z_{c,\text{med}} = 0.30, 0.30, 0.43$ , respectively.

## 2.2 Stellar mass estimation

We estimate the stellar masses,  $M_*$ , of red cluster member galaxies using the photometric data of the HSC-SSP Survey S19A (Aihara et al. 2019) and the SDSS Survey (DR16; Ahumada et al. 2020) as the supplementary photometric data. We first select red galaxies from the color-magnitude plane as a function of cluster redshift, following Nishizawa et al. (2018) and Okabe et al. (2019). We use the stellar population synthesis model of Bruzual & Charlot (2003) to estimate stellar masses from the Wide-layer depth grizy-band photometry at a given cluster redshift ( $z_c$ ). We adopt a single instantaneous burst at the formation redshift  $z = z_f$  (Oguri et al. 2018). We assume  $z_f = 3$  and the Chabrier IMF (Chabrier 2003). When we use the Salpeter IMF (Salpeter 1955), the masses are higher by a factor of  $\sim 1.5$  than those obtained with the Chabrier IMF (e.g. Pozzetti et al. 2007). We combine them with spectroscopically identified galaxies selected by a slice of  $|z - z_c| < 0.01(1 + z_c)$  from public spectroscopic redshifts in the HSC-SSP Survey region (Aguado et al. 2019; Skelton et al. 2014; Momcheva et al. 2016; Coil et al. 2011; Scodeggio et al. 2018). The HSC-SSP photometric data of some bright galaxies, such spectroscopically identified galaxies located in nearby clusters at  $z \sim 0.1$ , are too bright for the 8.2m Subaru telescope to be saturated (Aihara et al. 2018a, 2019). Some of them are flagged as saturated. We use complementary SDSS photometry for the missing galaxies. Moreover, we visually inspect whether large bright galaxies are missing in the catalog and add them if their visual properties are similar to those of galaxies in the catalog. The number of additional galaxies is only about twenty in the whole cluster sample. We use *cmodel* magnitudes (Lupton et al. 2001) from the HSC-SSP (Huang et al. 2018b; Bosch et al. 2018) and SDSS photometric data (Abazajian et al. 2004), which is

a linear-combination magnitude derived by the exponential and the de Vaucouleurs fits, and correct them with extinction. We confirm that the stellar masses estimated by the HSC and SDSS data agree with each other. The *cmodel* magnitude is a good total flux indicator to use as a universal magnitude for all types of objects (e.g. Lupton et al. 2001; Abazajian et al. 2004; Huang et al. 2018b; Bosch et al. 2018). However, it does not effectively include fluxes from the outer regions of massive galaxies, such as BCGs, where a diffuse intracluster light (ICL; e.g. Pillepich et al. 2018) is dominant and accounts for some fractions of the stellar mass (e.g. Huang et al. 2018a). We discuss this component in Sec. 4.1.5.

We then sum up stellar masses of galaxies within the projected, weak-lensing overdensity radii,  $r_{500}^{\text{WL}}$ , of individual clusters (Umetsu et al. 2020) from the XXL centers (Adami et al. 2018). We set the minimum stellar mass to be  $10^{10} M_{\odot}$ . Since the photometric data around bright stars are masked out, we correct the cylindrical, total stellar mass by the area fraction ( $F$ ) which is the ratio of the bright-star-masked area to total area within the overdensity radius. We here assume that the galaxies are uniformly distributed. We next subtract the stellar components associated with large-scale structure environment surrounding the targeting clusters and refer to them as the background component. The background component is estimated in an annulus between 2 Mpc and 4 Mpc to correct the projection effect;  $M_*^{\text{cyl}} = \sum_i M_{*i}(r_i < r_{500}^{\text{WL}})F - \sum_i M_{*i}(2 \text{ Mpc} < r_i < 4 \text{ Mpc})F_b$ , where  $i$  denotes the  $i$ -th galaxy within each region and the background component is also estimated by taking into account bright star mask corrections ( $F_b$ ). When we change the background annulus to 3-5 Mpc and 1.5-3.5 Mpc, the stellar masses change only by a few percent. We convert the cylindrical stellar masses to the spherical stellar masses by a deprojection using the NFW profile. We assume that the stellar mass density profile is described by the best-fit NFW mass density profile (Umetsu et al. 2020). In the deprojection method, we separate a central BCG from satellite galaxies, where the central BCG is defined by the largest stellar mass galaxy within 200 kpc from the XXL centers. We multiply the cylindrical stellar mass of the satellite galaxies,  $M_*^{\text{cyl}} - M_{\text{BCG}}$ , by a conversion factor,  $D_{\text{djp}}$ , which is obtained as the ratio between the spherical NFW mass within the measurement radius and an integration of the projected NFW profile out to the measurement radius. We use the concentration parameters for individual clusters in the computation of the conversion factor. The unweighted average of the concentration parameters is  $\langle c_{500}^{\text{WL}} \rangle = 2.8 \pm 1.5$ . When we change the concentration by  $\pm 1$ ,  $D_{\text{djp}}$  varies by only  $\pm 5$  percent. We then correct the obtained spherical stellar masses by a stellar mass function to consider the incompleteness of the stellar mass caused by the minimum cut of  $10^{10} M_{\odot}$ . We assume that the stellar mass function follows a Schechter luminosity function (Schechter 1976) and adopt the stellar mass function

of quiescent galaxies from the COSMOS Survey (Muzzin et al. 2013). We find that the correction factor is  $C \simeq 1$  and independent of cluster redshifts and we use the single value  $C = 1.0025$  for all the clusters. In short, the spherical mass estimate is described as

$$M_* = M_{\text{BCG}} + CD_{\text{dpj}}(M_*^{\text{cyl}} - M_{\text{BCG}}) = M_{\text{BCG}} + M_{\text{sat}}. \quad (1)$$

Here,  $M_{\text{sat}}$  is the spherical stellar mass of the satellite galaxies. We consider both the errors of the stellar mass of individual galaxies and the errors of weak-lensing overdensity radii. However, the errors due to weak-lensing overdensity radii ( $\sigma_*^{\text{err, WL}}$ ) account for more than 90 percent of the total error budget ( $\sigma_*^{\text{err}}$ ), and thus the other error sources ( $((\sigma_*^{\text{err}})^2 - (\sigma_*^{\text{err, WL}})^2)^{1/2}$ ) are negligible.

**Table 1.** Resulting regression parameters of the scaling relations between the cluster quantities ( $M_{500}^{\text{WL}}$ ,  $L_X$ ,  $M_*$ ,  $M_{\text{BCG}}$ , and  $M_g$ ) and the true mass  $M_{500}$  for the 136 XXL clusters. The normalization,  $\alpha$ , and the slope,  $\beta$ , are defined by the linear regressions (eq. E.2). The intrinsic scatter at a fixed true mass is represented by  $\sigma_{\text{int}}$ . <sup>†</sup>: the results using a trivariate Gaussian prior as the WL mass calibration, as described in Sec. 2.4.3. The errors denote the  $1\sigma$  uncertainty.

	$\alpha$	$\beta$	$\sigma_{\text{int}}$
$M_{500}^{\text{WL}} E(z)$	$-0.11_{-0.02}^{+0.02\dagger}$	$1.08_{-0.02}^{+0.02\dagger}$	$0.21_{-0.02}^{+0.02\dagger}$
$L_X E(z)^{-1}$	$0.29_{-0.13}^{+0.13}$	$1.38_{-0.18}^{+0.27}$	$0.73_{-0.14}^{+0.12}$
$M_* E(z)$	$0.76_{-0.08}^{+0.09}$	$0.85_{-0.09}^{+0.12}$	$0.52_{-0.06}^{+0.09}$
$M_{\text{BCG}} E(z)$	$-0.92_{-0.08}^{+0.08}$	$0.49_{-0.10}^{+0.11}$	$0.70_{-0.05}^{+0.06}$
$M_g E(z)$	$1.95_{-0.08}^{+0.08}$	$1.29_{-0.10}^{+0.16}$	$0.39_{-0.08}^{+0.08}$

### 2.3 Gas mass estimation

Gas masses within a fixed aperture of 500 kpc for all XXL clusters were published in DR2 (Adami et al. 2018). In the present analysis, we consider the gas mass measured within the weak-lensing overdensity radius ( $r_{500}^{\text{WL}}$ ). Here we provide updated gas masses from the *XMM-Newton* survey data for the 136 XXL clusters with available WL masses from HSC (Umetsu et al. 2020). Following Adami et al. (2018), we reduce the *XMM-Newton*/EPIC data using XMMSAS v13.5 and extract count images from all the available pointings in the [0.5 – 2] keV band. We use exposure maps for each observation to take the vignetting effect into account. We use a large collection of filter-wheel-closed data to extract models of the particle-induced background, and rescale the filter-wheel-closed data to match the count rates observed in the unexposed corners of each observation. To include all the available data, we create mosaic images by combining the count images, exposure maps, and background maps of all observations. To determine the

gas masses, we follow the method presented in Eckert et al. (2020) and implemented in the public code `pyproffit`<sup>1</sup>. For each cluster, we extract a surface brightness profile by accumulating source counts within concentric annuli around the cluster center as determined from the XXL detection pipeline XAMIN (Faccioli et al. 2018). We detect X-ray point sources by the XXL pipeline and mask them. The missing area is corrected to compute a surface brightness profile. The same procedure is applied to the background maps to create a model background count profile.

We model the three-dimensional gas emissivity profile as a combination of a large number of basis functions (King profiles) to allow a wide variety of shapes. The emissivity profile is projected along the line of sight and convolved with the instrumental PSF to predict the source brightness in each annulus. The residual sky background is fitted jointly to the data. The total model (source + background) is fitted to the data using the Hamiltonian Monte Carlo code PyMC3 (Salvatier et al. 2016). For more details on the reconstruction method, we refer the reader to Eckert et al. (2020).

To compute the conversion between emissivity and count rate, we simulate an absorbed APEC model (Smith et al. 2001) using XSPEC. Gas temperatures are taken from the spectral analysis of Giles et al. (2016) and the plasma is assumed to have a universal metal abundance of  $0.25Z_{\odot}$  (Anders & Grevesse 1989). We then relate the norm of the APEC model to the simulated source count rate,

$$K_{\text{APEC}} = \frac{10^{-14}}{4\pi(1+z)^2 d_A^2} \int n_e n_H dV \quad (2)$$

with  $n_e$ ,  $n_H$  the number density of electrons and ions, respectively. Finally, the gas mass within the WL overdensity radii,  $M_g \equiv M_{\text{gas}, 500}$ , is determined by integrating the reconstructed gas density profile within the weak-lensing  $r_{500}$ ,

$$M_g = \int_0^{r_{500}^{\text{WL}}} 4\pi r^2 \rho_{\text{gas}}(r) dr \quad (3)$$

with  $\rho_{\text{gas}} = \mu m_p (n_e + n_H)$  the gas mass density,  $\mu = 0.6$  the mean molecular weight, and  $m_p$  the proton mass. To propagate the uncertainties to the gas mass, for each set of posterior parameter values we integrate  $M_g$  using eq. 3 and randomize the value of the overdensity radius according to the posterior  $r_{500}^{\text{WL}}$  distribution (see Sec 2.4.1).

### 2.4 Multivariate scaling relations

We simultaneously estimate the multivariate scaling relations between weak-lensing mass ( $M_{500}^{\text{WL}}$ ), X-ray luminosity ( $L_X$ ), stellar mass ( $M_*$ ), BCG mass ( $M_{\text{BCG}}$ ), and gas mass ( $M_g$ ) by a Bayesian framework considering both selection effect and regression dilution bias. The details are described in Appendix A

<sup>1</sup> <https://github.com/domeckert/pyproffit>

**Table 2.** Intrinsic covariance for the 136 XXL clusters ( $L_X$ ,  $M_*$ ,  $M_{\text{BCG}}$ , and  $M_g$ ). The diagonal elements, the lower triangle elements, and the upper triangle elements represent express the intrinsic scatter ( $\sigma_{\ln Y_i}$ ), and a pair correlation coefficient ( $r_{ij}$ ), and an off-diagonal element of the intrinsic covariance ( $r_{i,j} \sigma_{\ln Y_i} \sigma_{\ln Y_j}$ ), respectively. The errors and the lower bound denote the  $1\sigma$  uncertainty and the  $1\sigma$  lower limit, respectively.

	$L_X E(z)^{-1}$	$M_* E(z)$	$M_{\text{BCG}} E(z)$	$M_g E(z)$
$L_X E(z)^{-1}$	$0.73^{+0.12}_{-0.14}$	$0.07^{+0.14}_{-0.09}$	$0.08^{+0.10}_{-0.08}$	$0.28^{+0.12}_{-0.14}$
$M_* E(z)$	$0.20^{+0.23}_{-0.28}$	$0.52^{+0.09}_{-0.06}$	$0.24^{+0.07}_{-0.06}$	$0.04^{+0.10}_{-0.05}$
$M_{\text{BCG}} E(z)$	$0.18^{+0.14}_{-0.19}$	$0.67^{+0.06}_{-0.09}$	$0.70^{+0.06}_{-0.05}$	$0.03^{+0.06}_{-0.05}$
$M_g E(z)$	$> 0.97$	$0.24^{+0.31}_{-0.33}$	$0.16^{+0.17}_{-0.22}$	$0.39^{+0.08}_{-0.08}$

and B. For the description of this approach see Sereno (2016) and Sereno et al. (2020). We use the natural logarithm ( $\ln$ ) for observables and their intrinsic scatter. We adopt the Markov Chain Monte Carlo (MCMC) method. XXL candidate clusters are classified as C1 or C2 according to total count rate and core size to avoid contamination of point sources (Pacaud et al. 2006, 2016). The C1 selection can be assimilated to a surface brightness limit, while the C2 subsample used in this paper is unconstrained (Sec. 2.1).

We approximately use the X-ray luminosity  $L_X$  in the [0.5–2] keV band as a simple selection function instead of the two parameters of the total count rate and core size, following Sereno et al. (2020). The DR2 catalog (Adami et al. 2018) includes X-ray flux within 60 arcsec from the XXL centers and X-ray luminosity within  $r_{500}$  determined by a mass and temperature scaling relation based on a CFHTLenS shape catalog (Lieu et al. 2016). However, out of 136 clusters, the public X-ray luminosities of 32 clusters are not available, which does not meet the sample size of this paper. Furthermore, to avoid the systematics inherent in the mass scaling relation and WL mass measurement, we remeasure the total X-ray luminosity in the [0.5–2] keV. We first compute the maximum detection radius ( $r_{\text{max}}$ ) of each source as the radius at which the reconstructed surface brightness is 10% of the locally determined background brightness. We then integrate the surface brightness within the corresponding circular area to measure the total count rate. Finally, we use XSPEC to compute the conversion between count rate and luminosity at the redshift of the source assuming the source spectrum is described by an absorbed APEC model, and then we obtain the total luminosity  $L_X$  which is integrated within the maximum detection radius  $r_{\text{max}}$ . Here, it is important that the measurement of the X-ray luminosity does not use external information but X-ray data alone. The maximum radius  $r_{\text{max}}$  is positively correlated with  $r_{500}^{\text{WL}}$  and the scatter is 0.11 dex. When we remove 19 large-offset clusters of which  $r_{\text{max}}$  and  $r_{500}^{\text{WL}}$  differ by more than twice as large as  $r_{500}^{\text{WL}}$  errors, we find that the results do not change significantly. This is caused by no error correlation between  $r_{\text{max}}$  and  $r_{500}^{\text{WL}}$ . We approximately use  $L_X E(z)^{-1}$  as expected by a self-similar solution because

the measurement radii sufficiently covers the X-ray dominated region, where  $E(z) = (\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0})^{1/2}$ . We employ the minimum X-ray luminosity as the threshold in the regression analysis (eq. A2). Since the measurement radius of  $L_X$  is independent of the WL overdensity radius, there is no error correlation with other quantities. We introduce the natural logarithmic quantities for the observables in Bayesian inference, defined as

$$x = \ln \frac{M_{500}^{\text{WL}} E(z)}{10^{14} M_{\odot}}, \quad (4)$$

$$\mathbf{y} = \left\{ \ln \frac{L_X E(z)^{-1}}{10^{43} \text{ ergs}^{-1}}, \ln \frac{M_* E(z)}{10^{12} M_{\odot}}, \ln \frac{M_{\text{BCG}} E(z)}{10^{12} M_{\odot}}, \ln \frac{M_g E(z)}{10^{12} M_{\odot}} \right\}.$$

We assume the  $E(z)$  dependence expected by the self-similar solution for all the observables. We refer to baryonic observables as  $\mathbf{y} = \{y_l, y_*, y_{\text{BCG}}, y_g\}$ . We express the true values of their observed quantities as capital letters,  $X$  and  $\mathbf{Y}$ . The observables and true variables are related by  $p(x, \mathbf{y} | X, \mathbf{Y}) = \mathcal{N}(\{x, \mathbf{y}\} | \{X, \mathbf{Y}\}, \Sigma_{\text{err}})$ , where  $\mathcal{N}$  and  $\Sigma_{\text{err}}$  denote a normal distribution and an observational error covariance matrix, respectively (Appendix A). We aim at measuring the linear regression with respect to the actual quantities of  $X$  and  $\mathbf{Y}$  (Sec. 2.4.2). The observational errors of  $x$  and  $\mathbf{y}$  are described by fractional errors  $\sigma_x^{\text{err}} = \sigma_{\text{WL}}^{\text{err}}$  and  $\sigma_{\mathbf{y}}^{\text{err}} = \{\sigma_l^{\text{err}}, \sigma_*^{\text{err}}, \sigma_{\text{BCG}}^{\text{err}}, \sigma_g^{\text{err}}\}$ . The diagonal elements in  $\Sigma_{\text{err}}$  are  $\{(\sigma_x^{\text{err}})^2, (\sigma_{\mathbf{y}}^{\text{err}})^2\}$ . The details of the error covariance matrix are described in Sec. 2.4.1.

Redshift dependence in eq. (4) can be obtained for self-similar evolution. Since the critical density  $\rho_{\text{cr}}(z) \propto H(z)^2 \propto E(z)^2$  and the overdensity radius  $r_{500} \propto c/H(z) \propto E(z)^{-1}$ , the mass becomes  $M_{500} \propto \rho_{\text{cr}}(z) r_{500}^3 \propto E(z)^{-1}$ . Here,  $H(z)$  is the Hubble parameter at given redshift and  $c$  is the light velocity. Thus,  $M_{500} E(z)$  is independent of redshift. Assuming that the baryonic mass density is proportional to the critical density, that is, a constant baryon fraction against both mass and redshift, we similarly obtain  $M_i E(z)$  where  $i = \{*, \text{BCG}, g\}$ . Since the soft-band X-ray luminosity is proportional to the square of the electron number density ( $n_e$ ) in the volume,  $L_X \propto n_e^2 (c/H(z))^3 \propto E(z)$  where  $n_e \propto \rho_{\text{cr}}(z)$ . Therefore,  $L_X E(z)^{-1}$  becomes constant against redshift. We show the result without the  $E(z)$  dependence in Appendix E.

### 2.4.1 Observational error covariance matrix

The relationship between the actual ( $X$  and  $Y$ ) and observed ( $x$  and  $y$ ) quantities is expressed by a multivariate Gaussian distribution with an observational error covariance matrix,  $\Sigma_{\text{err}}$ . The diagonal elements of the error covariance matrix consists of the variance of the observational errors. The error correlation in the off-diagonal elements of the error covariance matrix is expressed by the subscript combinations of  $r_{i,j}^{\text{err}} \sigma_i^{\text{err}} \sigma_j^{\text{err}}$  between the  $i$  and  $j$  components, where  $r_{i,j}^{\text{err}}$  is the error correlation coefficient with  $i, j = \{\text{WL}, l, *, \text{BCG}, g\}$ . The error correlation coefficient,  $r_{i,j}^{\text{err}}$ , describes an error propagation by the same measurement radii of the weak-lensing, stellar, and gas masses because the stellar and gas masses are computed within the spherical overdensity radii ( $r_{500}^{\text{WL}}$ ) of the weak-lensing masses (Umetsu et al. 2020).

We first estimate the error correlation coefficient,  $r_{\text{WL},*}^{\text{err}}$ , between the weak-lensing mass and the stellar mass. Since member galaxies are sparsely distributed, the errors of the weak-lensing masses randomly affect the stellar mass estimations in the individual clusters and it is difficult to independently measure individual error correlations. We therefore use the same correlation coefficient,  $r_{\text{WL},*}^{\text{err}}$ , estimated by the whole sample of clusters. We randomly pick up the overdensity radius by drawing values according to the posterior distributions of weak-lensing overdensity radii for the individual clusters. We correspondingly measure stellar masses within the given radii and then evaluate the error correlation coefficient by combining all the clusters. The number of realizations is 1000 for each cluster. We find  $r_{\text{WL},*}^{\text{err}} = 0.873$ . The error correlation between  $M_{\text{BCG}}$  and  $M_*$  is negligible because the error of  $M_*$  is mainly due to the weak-lensing overdensity radii.

Since the gas mass density is smoothly distributed, we easily obtain the correlation coefficient of measurement error  $r_{\text{WL},g}^{\text{err}}$  between the weak-lensing mass and the gas mass for individual clusters. We employ the same method as the estimation of  $r_{\text{WL},*}^{\text{err}}$ . The average correlation coefficient for the whole sample is  $\langle r_{\text{WL},g}^{\text{err}} \rangle = 0.908$ .

The measurement errors for the stellar mass and the gas mass are also correlated through the same overdensity radii. It is, however, difficult to measure  $r_{*,g}^{\text{err}}$  for individual clusters because the sparse distribution of the member galaxies makes it difficult to estimate individual error correlation coefficients. We therefore estimate  $r_{*,g}^{\text{err}}$  by a trigonometric formula (e.g. Rousseeuw & Molenberghs 1994) which is derived from the definition that the determinant of covariance matrix must be positive. The expected range of  $r_{*,g}^{\text{err}}$  is  $r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}} - \sqrt{1 - (r_{\text{WL},*}^{\text{err}})^2} \sqrt{1 - (r_{\text{WL},g}^{\text{err}})^2} \leq r_{*,g}^{\text{err}} \leq r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}} + \sqrt{1 - (r_{\text{WL},*}^{\text{err}})^2} \sqrt{1 - (r_{\text{WL},g}^{\text{err}})^2}$ . When we uniformly and randomly pick up values within the ranges, the average result approximates mean value,  $r_{*,g}^{\text{err}} = r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}}$ . We therefore adopt the mean value,  $r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}}$ , for individual

clusters to derive the regression parameters, and then incorporate the results with the lower and upper bounds of the trigonometric formula into the parameter errors. The average value for whole sample is  $\langle r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}} \rangle = 0.793$ . The lower and upper ranges of  $r_{*,g}^{\text{err}}$  increase the measurement error of the intrinsic correlation coefficient between  $M_*$  and  $M_g$  by  $\sim 100\%$ , while the other parameters are insensitive. If we set  $r_{*,g} = 0$ , an acceptance ratio of the MCMC chain becomes close to zero and the parameters cannot be constrained because it does not satisfy with the condition of the error correlation matrix.

### 2.4.2 Linear regression and intrinsic covariance

The linear regressions between the actual quantities ( $X$  and  $Y$ ) and a true mass ( $Z$ ) are described by

$$X_Z = \alpha_X + \beta_X Z, \quad (5)$$

$$Y_Z = \alpha + \beta Z, \quad (6)$$

where  $Z = \ln(M_{500} E(z) / 10^{14} M_{\odot})$  is the logarithmic value of the true mass,  $\alpha_X$  and  $\alpha$  are the normalizations, and  $\beta_X$  and  $\beta$  are the slopes of the mass dependence. We consider the intrinsic covariance matrix,  $\Sigma_{\text{int}}$  (Okabe et al. 2010), which describes the statistical properties of cluster baryonic components. The diagonal elements of the intrinsic covariance are specified by fractional scatter  $\sigma_X^2 = \sigma_{\text{WL}}^2$  and  $\sigma_Y^2 = \{\sigma_l^2, \sigma_*^2, \sigma_{\text{BCG}}^2, \sigma_g^2\}$ . The intrinsic correlation coefficient in the off-diagonal elements is expressed as the correlation  $r_{i,j}$  between the  $i$  and  $j$  components, where  $i, j = \{l, *, \text{BCG}, g\}$ . Since it is difficult to constrain the intrinsic correlation coefficient associated with the weak-lensing mass, we fix  $r_{\text{WL},i} = 0$ . We use flat prior for the parameters.

We also consider a single Gaussian distribution of  $p(Z)$  as a parent population of  $Z$  in the Bayesian analysis in order to correct both regression dilution effect and selection effect (Appendix A). The parent population  $p(Z)$  is simultaneously determined by the scaling relation between the total X-ray luminosity and the mass, where the X-ray luminosity is approximately the tracer of the cluster finder. It also can be determined by weak-lensing masses with the mass calibration, which is discussed in Sec 4.1.2. Due to the cosmological dimming of X-ray emission, we expect that more massive clusters can be found at higher redshift (Sereno et al. 2020). We therefore introduce a redshift dependence of the parent population,  $p(Z, z) = \mathcal{N}(\mu_Z(z), \sigma_Z(z))$ , of which the mean and standard deviation are described by

$$\mu_Z(z) = \mu_{Z,0} + \gamma_{\mu_Z} \ln E(z), \quad (7)$$

$$\sigma_Z(z) = \sigma_{Z,0} E(z)^{\gamma_{\sigma_Z}}, \quad (8)$$

where  $\gamma_{\mu_Z}$  and  $\gamma_{\sigma_Z}$  are the redshift dependence of the mean and standard deviation, respectively. The parameters of  $p(Z, z)$  are non-informative, hyper-parameters and simultaneously derived by the Bayesian analysis. Thus, the result of multivariate scal-

ing relations is independent of the cluster number counts and of the cosmological parameters. We note that  $\sigma_Z(z)$  is important to accurately determine the slopes by considering the regression dilution effect (Appendix A and B). Although we tried to fit with double Gaussian distributions of the  $Z$  distribution, we could not constrain the parameters of the second Gaussian component. Thus, the single  $p(Z, z)$  is sufficient for this analysis. Other possibilities including no-redshift evolution of  $p(Z)$  will be discussed in Sec. 4.1.2.

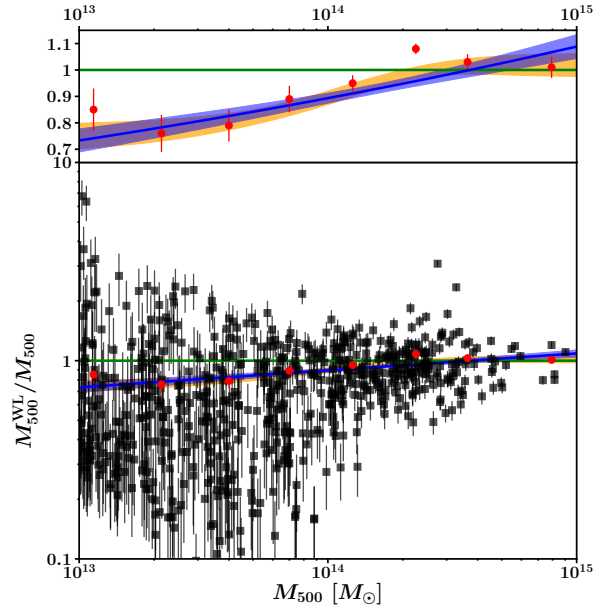
#### 2.4.3 Weak-lensing mass calibration

The parameters,  $\alpha_{\text{WL}}$ ,  $\beta_{\text{WL}}$ , and  $\sigma_{\text{WL}}$ , describe our weak-lensing mass calibration. Weak-lensing mass estimates for individual clusters are scattered from their true values because of their non-spherical halo shape, substructures, and surrounding large-scale structure (e.g. Hoekstra 2003; Becker & Kravtsov 2011; Oguri & Hamana 2011; Okabe et al. 2016; Umetsu 2020). Moreover, even when averaged over many clusters, their ensemble mass estimates can be biased, if the true mass profiles deviate from the assumed profile (e.g. Umetsu et al. 2020). Umetsu et al. (2020) validated their weak-lensing mass estimates for cluster and group scales using both cosmological numerical simulations (McCarthy et al. 2017, 2018) and analytical NFW models, and found that the weak lensing mass bias weakly depends on true masses. In our multivariate regression analysis, we consider the bias and the scatter between the weak-lensing mass,  $M_{500}^{\text{WL}}$ , and the true mass,  $M_{500}$ .

Umetsu et al. (2020) and Sereno et al. (2020) only accounted for the  $\pm 5\%$  calibration uncertainty due primarily to observational systematics in their observable–mass scaling relations. In the mass forecasting for the  $M_{500} - T_X$  relation of Umetsu et al. (2020), they applied an additional constant mass-modeling bias correction of  $-11\%$  evaluated at the mean mass scale of the XXL sample.

We characterize the mass dependence of the NFW weak-lensing mass estimates, or the  $M_{500}^{\text{WL}} - M_{500}$  relation, using the results of Umetsu et al. (2020) based on synthetic weak-lensing observations of 639 cluster halos in the dark-matter-only run of BAHAMAS simulations (McCarthy et al. 2017). As shown in Figure 1, the mass bias increases with true mass in the regime of low masses and it is nearly constant in the high mass range ( $M_{500} \gtrsim 10^{14} M_\odot/h$ ). This can be approximated with a tanh functional form. We fit the data with a functional form of  $M_{500}^{\text{WL}}/M_{500} = a \tanh M_{500}/b + c$  and the intrinsic scatter of  $\sigma_{\text{WL}}$ . We find  $a = 0.29 \pm 0.06$ ,  $b = 1.57^{+0.97}_{-0.48}$ ,  $c = 0.73^{+0.06}_{-0.06}$ , and  $\sigma_{\text{WL}} = 0.21 \pm 0.02$  (orange region in Figure 1).

However, the mathematical formulation in the regression analysis (Appendix A) requires a power-law relation between the true mass and weak-lensing mass, or a linear relation between their logarithmic quantities. We here assume eq. (5) and find  $\alpha_{\text{WL}} = \ln(0.89 \pm 0.02)$ ,  $\beta_{\text{WL}} = 1.09 \pm 0.02$  and  $\sigma_{\text{WL}} =$



**Fig. 1.** Weak-lensing mass calibration. Red circles denote the the average mass bias at each mass bin from Umetsu et al. (2020). The solid blue line and regions denote the estimate and  $1\sigma$  uncertainty for the power-law relation ( $\ln M_{500}^{\text{WL}} = \alpha_{\text{WL}} + \beta_{\text{WL}} \ln M_{500}$ ), respectively. The orange region is the  $1\sigma$  uncertainty for the function of  $M_{500}^{\text{WL}}/M_{500} = a \tanh M_{500}/b + c$ . The green horizontal line expresses no bias. The black squares are the results of simulated clusters (Umetsu et al. 2020). The top and bottom panels display data and result on a linear scale and a logarithmic scale, respectively. The tanh form better describes the data.

$0.21 \pm 0.02$  for the mean relationship between the weak-lensing mass and the true mass, as represented by the blue line in Figure 1. The result agrees with that of the tanh function within the  $1\sigma$  uncertainty at  $M_{500} E(z) \lesssim 1.5 \times 10^{15} M_\odot$ .

We use a trivariate Gaussian distribution of  $\mathcal{N}_{3\text{D}}(\alpha_{\text{WL}}, \beta_{\text{WL}}, \ln \sigma_{\text{WL}})$  as a prior for the weak-lensing mass calibration. The covariance matrix in  $\mathcal{N}_{3\text{D}}$  employs the error covariance matrix of the linear regression in the power-law mass calibration. Therefore, all the mass calibration uncertainties are propagated into the results. We discuss a case of the tanh function in Sec. 4.1.3.

## 3 Results

### 3.1 Normalization and slopes of scaling relations

The Bayesian framework straightforwardly derives the normalization, slopes, and intrinsic covariance of the multivariate scaling relations. The number of parameters is 25 including 4 hyper-parameters. We use biweight estimates of marginalized posterior distributions as the parameter estimates.

The estimated normalizations and slopes for the  $L_X - M$ ,  $M_g - M$ ,  $M_* - M$ , and  $M_{\text{BCG}} - M$  relations are shown in Table 1. The posterior distribution is shown in Appendix D (Figure 15). The slope of the X-ray luminosity,  $1.38^{+0.27}_{-0.18}$ ,

is  $2.1\sigma$  higher than the prediction of the self-similar model ( $\beta = 1$ ). Figures 2, 3, and 4 show the resulting scaling relations of the gas, stellar, and BCG masses with the weak-lensing masses ( $M_{500}^{\text{WL}}$ ; Umetsu et al. 2020), respectively. The scaling relations shown in the figures are described by  $Y_Z = \alpha - (\beta/\beta_{\text{WL}})\alpha_{\text{WL}} + (\beta/\beta_{\text{WL}})X_Z$ . For comparison, we also plot the direct observables and the stacked observables sorted by the X-ray luminosity. We did not fit using the stacked observables. We stack 18 clusters in each subsample in ascending order of the C1/C2 X-ray luminosity, and the numbers of the remaining C1 and C2 clusters in the highest luminosity subsamples are 11 and 17, respectively. Since the stacked quantities are sorted by the X-ray luminosity, the subsample grouping is independent of any observables in the  $x$ -axis of Figures 2-4 or the  $y$ -axis of Figures 3 and 4. The stacked quantities are computed by

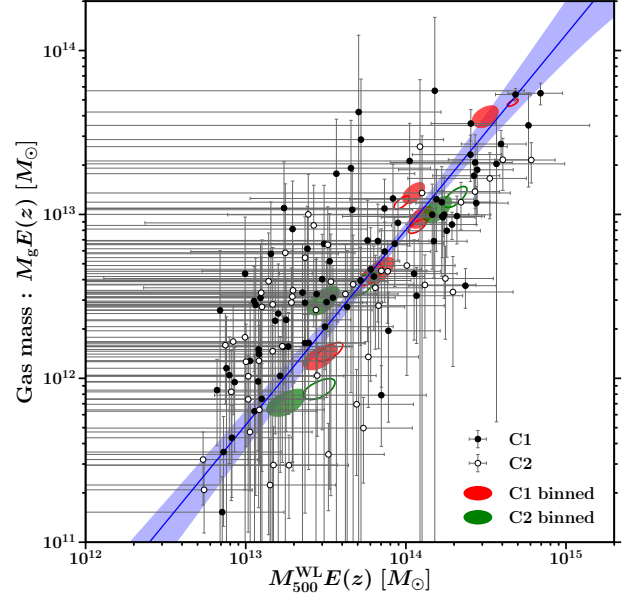
$$\langle \mathbf{v} \rangle = \left( \sum_n (\Sigma_n)^{-1} \right)^{-1} \sum_n (\Sigma_n)^{-1} \mathbf{v}_n \quad (9)$$

where  $\mathbf{v} = \{x, y\}$ ,  $\Sigma$  is the error matrix,  $\Sigma_{\text{err}}$ , or the composition matrix of  $\Sigma_{\text{err}} + \Sigma_{\text{int}}$  for  $\mathbf{v}$ , and  $n$  is the  $n$ -th cluster. The mean observables, weighted with the error matrix, show some scatter around the scaling-relation baselines, which exhibits intrinsic scatter. Such a feature is visible especially in the  $M_{\text{BCG}} - M$  relation with the largest intrinsic scatter. We thus weight them with the composition matrix to compare with the baselines shown in blue in Figures 2-4, and find that the stacked quantities are in good agreement with the baselines. It also indicates the consistency of Bayesian inference among  $\alpha$ ,  $\beta$ , and  $\Sigma_{\text{int}}$  to explain the data.

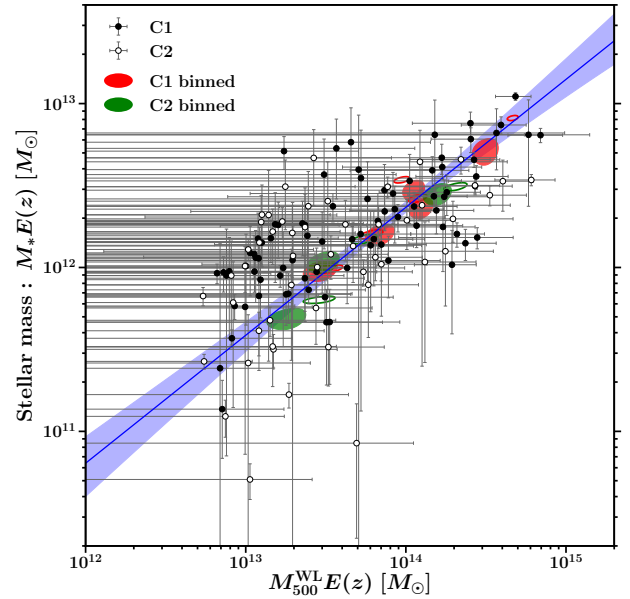
We find that the slopes in the  $M_g - M_{500}$  and  $M_* - M_{500}$  relations are,  $1.29^{+0.16}_{-0.10}$ , and  $0.85^{+0.12}_{-0.09}$ , steeper and shallower than the self-similar predictions ( $\beta = 1$ ), respectively. The significance levels of the deviations from unity are  $\sim 3\sigma$  and  $1.5\sigma$ , respectively. We find a shallower slope,  $\beta_{\text{BCG}} = 0.49^{+0.11}_{-0.10}$ , in the  $M_{\text{BCG}} - M_{500}$  relation, which indicates that the BCG stellar mass has only a weak dependence on the halo mass.

### 3.2 Parent Population

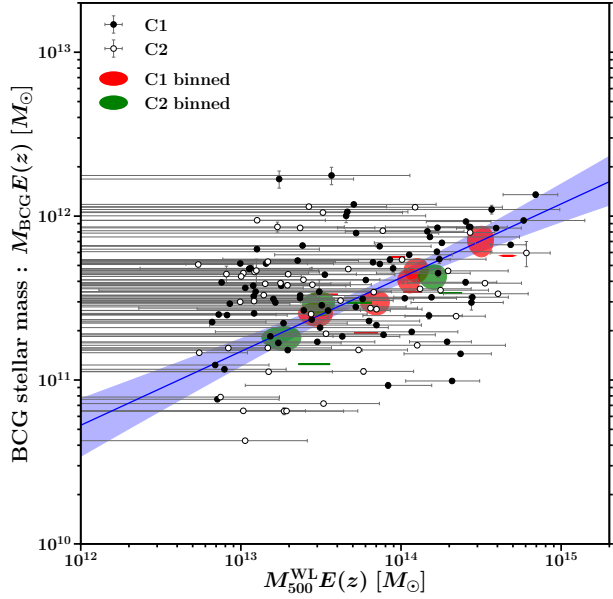
The resulting regression parameters for the parent population are  $\mu_{Z,0} = -1.02^{+0.20}_{-0.20}$ ,  $\gamma_{\mu_Z} = 3.53^{+0.68}_{-0.62}$ ,  $\sigma_{Z,0} = 1.21^{+0.26}_{-0.24}$ , and  $\gamma_{\sigma_Z} = -2.38^{+0.82}_{-1.00}$ . The mean mass and the standard deviation of the parent population increases and decreases with increasing redshift, respectively. Thus, the more massive clusters at higher redshifts are discovered by the XXL Survey, as expected due to the X-ray dimming effect. The mass distributions at lower redshift is broader than those at higher redshift, indicating that it is easier to find clusters from a broad mass range at lower redshift.



**Fig. 2.** The gas mass and weak-lensing mass relation. The solid blue line and region denote the estimate and  $1\sigma$  uncertainty for the scaling-relation baseline, respectively. The black and white circles are the C1 and C2 subsamples, respectively. The red and green ellipses represent the 68 percent confidence levels for the stacked quantities of the subsamples of the C1 and C2 clusters which are sorted by the X-ray luminosity, respectively. Filled and open ellipses are those weighted with the covariance matrix  $\Sigma = \Sigma_{\text{err}} + \Sigma_{\text{int}}$  and  $\Sigma_{\text{err}}$ , respectively. The stacked quantities follow the resulting baseline well.



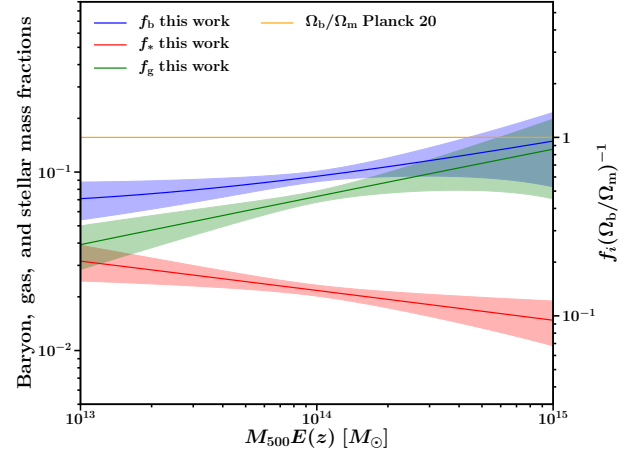
**Fig. 3.** The stellar mass and weak-lensing mass relation. The solid blue line and region denote the estimate and  $1\sigma$  uncertainty for the scaling-relation baseline, respectively. The black and white circles and the red and green ellipses have the same meaning as Figure 2.



**Fig. 4.** The BCG mass and weak-lensing mass relation. The solid blue line and region denote the estimate and  $1\sigma$  uncertainty for the scaling-relation baseline, respectively. The black and white circles and the red and green ellipses have the same meaning as Figure 2. Since the measurement errors of the BCG mass are much smaller than those of the WL masses, the shapes of the stacked quantities weighted with  $\Sigma_{\text{err}}$  become lines along the  $x$  axis.

### 3.3 Baryon fractions

We convert the resulting scaling relations to the baryon, gas and stellar fractions as a function of the true halo mass ( $M_{500}$ ) not the weak-lensing mass ( $M_{500}^{\text{WL}}$ ),  $f_i = M_i(< r_{500})/M_{500}(< r_{500})$ , where  $i = \{b, g, *\}$  and  $r_{500}$  is the overdensity radius of the true mass. Since the baselines,  $Y_Z(< r_{500}^{\text{WL}})$ , are computed by using  $M_g(< r_{500}^{\text{WL}})$  and  $M_*(< r_{500}^{\text{WL}})$  measured within the WL overdensity radii  $r_{500}^{\text{WL}}$ , we convert to those measured within  $r_{500}$ . The details are described in Appendix C. The aperture correction depends on the baryonic mass density slope, the mass calibration, and the true mass (eq. A16). As for the stellar mass profile, we assumed that the stellar mass density profile follows the dark matter profile with the average concentration parameter  $\langle c_{500}^{\text{WL}} \rangle = 2.8$  (Sec. 2.2). We assumed the King model of the electron number density follows  $n_e \propto r^{-3\beta}$  with  $\beta = 2/3$  outside gas cores (Sec. 2.3). The stellar mass normalization with the aperture correction becomes  $\sim 1.05$ ,  $\sim 1.02$ , and  $\sim 0.99$  times that without the correction at  $M_{500}E(z) \sim 10^{13} M_\odot$ ,  $\sim 10^{14} M_\odot$ , and  $\sim 10^{15} M_\odot$ , respectively. As for the gas mass, the aperture correction changes the normalization by  $\sim 1.09$ ,  $\sim 1.04$ , and  $\sim 0.97$  times at  $M_{500}E(z) \sim 10^{13} M_\odot$ ,  $\sim 10^{14} M_\odot$ , and  $\sim 10^{15} M_\odot$ , respectively. Figure 5 shows the resulting baryon, gas and stellar fractions. Since the power-law mass calibration is validated in the true-mass range of  $10^{13} M_\odot \lesssim M_{500}E(z) \lesssim 10^{15} M_\odot$ , the lower and upper bounds of the  $x$ -axis in Figure 5 are set to be  $10^{13} M_\odot$  and  $10^{15} M_\odot$ , respectively. It fully covers the true mass population at  $0 \lesssim z \lesssim 1$ .



**Fig. 5.** Baryon (blue), gas (green), and stellar (red) mass fractions as a function of the halo mass. The shade regions are the  $1\sigma$  uncertainty. The orange horizontal line is  $\Omega_b/\Omega_m$  (Planck Collaboration et al. 2020).

We do not show the observables in the same figure because the quantity in the  $x$ -axis is not the weak-lensing mass but the true mass. Although the true masses can be statistically calculated by the mean relationship of the weak-lensing mass calibration (eq. 5) and its intrinsic scatter, an actual weak-lensing mass bias or true mass of each cluster is unclear.

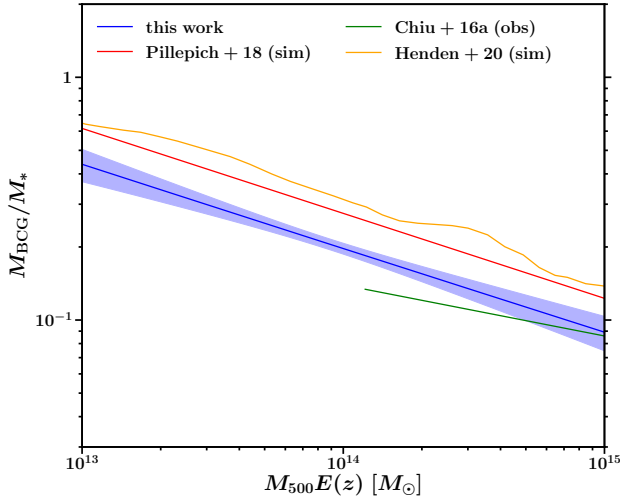
The uncertainties in Figure 5 fully take into account the error covariance matrix of the linear regressions. The gas mass fraction,  $f_g = M_g(< r_{500})/M_{500}$ , increases as the halo mass increases, reaching 90 percent of  $\Omega_b/\Omega_m = 0.1564 \pm 0.0016$  (Planck Collaboration et al. 2020) at  $M_{500}E(z) \sim 10^{15} M_\odot$ . In contrast, the stellar mass fraction,  $f_* = M_*(< r_{500})/M_{500}$ , decreases as the halo mass increases. These trends are the same as the slope deviations from unity in the scaling relations (Sec. 3.1).

The total baryon mass fraction,  $f_b = f_g + f_*$ , is  $\sim 50$  percent of  $\Omega_b/\Omega_m$  at  $M_{500}E(z) \sim 10^{13} M_\odot$ ,  $\sim 60$  percent at  $\sim 10^{14} M_\odot$ , and  $\sim 100\%$  at  $\sim 10^{15} M_\odot$ . The mass-dependent slope of  $f_b$  on group scales of  $\lesssim 10^{14} M_\odot$  is less steep than that on massive clusters of  $\gtrsim 5 \times 10^{14} M_\odot$ .

When we use the Salpeter IMF, the baryon and stellar mass fractions at  $\sim 10^{13} M_\odot$  are  $\sim 1.2$  and  $\sim 1.5$  times higher than those derived by the Chabrier IMF, respectively. At  $\sim 10^{14} M_\odot$ , the baryon fraction increases only by  $\sim 1.1$  times. The overall trends do not significantly change by a choice of the IMF.

### 3.4 BCG mass to total stellar mass ratio

We compute the BCG stellar mass to total stellar mass ratio as a function of the true mass (Figure 6) from the  $M_{\text{BCG}} - M$  and  $M_* - M$  relations. Since the BCG stellar mass measurement is independent of the weak-lensing overdensity radius, it is independent of the aperture correction. Since the errors of the linear regressions are correlated with each other, the error covariance



**Fig. 6.** The BCG mass to total stellar mass ratio as a function of halo mass. The solid red and orange lines are the simulations of Pillepich et al. (2018) and Henden et al. (2020), respectively. We recompute  $M_{\text{BCG}}/M_*$  from the  $M_* - M$  and  $M_{\text{BCG}} - M$  lines (Chiu et al. 2016a), shown as the solid green line.

matrix is taken into account to compute the errors of the ratio. The fraction of the BCG in the total mass is at most  $\sim 10$  percent at  $M_{500}E(z) \sim 10^{15} M_\odot$  and  $\sim 20$  percent at  $\sim 10^{14} M_\odot$ . However, the fraction at  $\sim 10^{13} M_\odot$  accounts for  $\sim 45$  percent. Therefore, the BCG is a more dominant component of the stellar mass components on a group scale.

### 3.5 Intrinsic covariance of baryon contents

Another important property of the multivariate scaling relations is the intrinsic covariance. Table 2 describes the resulting intrinsic covariance (see also Table 1). The diagonal element, the lower off-diagonal element, and the upper off-diagonal element are intrinsic scatter, a pair correlation coefficient, and an off-diagonal element of the intrinsic covariance at fixed cluster mass, respectively. The posterior distribution is shown in Appendix D (Figure 16).

The intrinsic scatter in gas mass is  $\sigma_g = 0.39^{+0.08}_{-0.08}$ , corresponding to  $\sim 0.17$  dex. The intrinsic scatter in the stellar mass,  $\sigma_* = 0.52^{+0.09}_{-0.06} \sim 0.23$  dex, is larger than  $\sigma_g$ . The intrinsic scatter of the BCG mass is  $\sigma_{\text{BCG}} = 0.70^{+0.06}_{-0.05} \sim 0.30$  dex. The largest intrinsic scatter is in BCG stellar mass, followed in order by stellar mass and gas mass;  $\sigma_g < \sigma_* < \sigma_{\text{BCG}}$ . The scatter trend is visually confirmed in Figures 2-4. The error-covariance-weighted means for the subsamples binned by the X-ray luminosity show some scatter in the scaling-relations (Figures 2-4). Comparisons of numerical simulations and other observations are discussed in Sec. 4.

We find strong intrinsic correlation coefficients between stellar mass and BCG mass and between X-ray luminosity and gas mass;  $r_{*,\text{BCG}} = 0.70^{+0.06}_{-0.05}$  and  $r_{l,g} > 0.97$ . Other intrinsic cor-

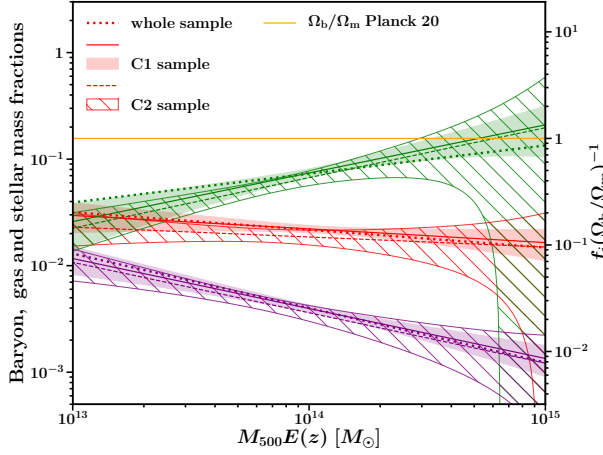
relation coefficients agree with no correlation within the errors;  $r_{*,g} = 0.24^{+0.31}_{-0.33}$ , and  $r_{\text{BCG},g} = 0.16^{+0.17}_{-0.22}$ . We also explored the possibility that a spurious positive or negative correlation could be caused by a finite sampling size. We assess an accidental probability that 136 random pairs give the observed intrinsic correlation coefficient, following Okabe et al. (2010). The accidental probability,  $\mathcal{P}(r \geq |r_{i,j}|)$ , is specifically defined as follows; the correlation coefficient of the two random variables in a sample of 136 drawings is higher than the absolute value of the intrinsic correlation coefficient. It corresponds to a probability of the null hypothesis that the two correlation is zero. The resulting maximum  $p$ -values are  $\mathcal{P}(r \geq |r_{*,\text{BCG}}|) \sim \mathcal{O}(10^{-13})$ ,  $\mathcal{P}(r \geq |r_{*,g}|) \sim 3 \times 10^{-1}$ ,  $\mathcal{P}(r \geq |r_{\text{BCG},g}|) \sim 4 \times 10^{-1}$ , and  $\mathcal{P}(r \geq |r_{l,g}|) \sim 0$ , respectively. We therefore reject a possibility of the accidental correlation between  $M_{\text{BCG}}$  and  $M_*$  and between  $L_X$  and  $M_g$ .

We also study the intrinsic correlation coefficient between BCG stellar mass and satellite galaxy mass defined by  $M_{\text{sat}} = M_* - M_{\text{BCG}}$ . The intrinsic correlation coefficient,  $r_{\text{sat},\text{BCG}} = 0.51^{+0.10}_{-0.16}$ , is significant. We find that the intrinsic correlation coefficient between  $M_{\text{sat}}$  and  $M_g$  is consistent with no correlation;  $r_{\text{sat},g} = 0.43^{+0.30}_{-0.59}$ .

### 3.6 C1 and C2 subsamples

Our sample comprises 83 C1 and 53 C2 clusters from the XXL DR2 sample (Adami et al. 2018). We here split the whole sample into the C1 and C2 subsamples. Since the mean luminosity for the C2 sample is lower than that for the C1 sample (Adami et al. 2018), we set the maximum threshold in  $L_X$  as the highest X-ray luminosity among the C2 sample in the Bayesian analysis. Even when we remove the upper bound of the X-ray luminosity, the results do not significantly change. The resulting regression parameters for the C1 and C2 subsamples are shown in top panel of Table 3. The baryon fractions for the whole sample agree with those for the C1 and C2 samples (Figure 7), except for the low-mass end of the  $M_g - M$  relation. The gas fractions between the whole sample and the C1 sample and between the whole sample and the C2 sample differ by  $\sim 1.1\sigma$  and  $\sim 1.5\sigma$  at  $10^{13} M_\odot$ , respectively. This small discrepancy is caused by the steeper C1 and C2 slopes of the  $M_g - M$  relation. When we fix the unconstrained intrinsic scatter,  $\sigma_g$ , for the C1 and C2 samples with  $\sigma_g = 0.39$  obtained by the whole sample, we find that they agree within  $1\sigma$ . The determination of the slopes is associated with the intrinsic scatter.

The intrinsic covariances for the two subsamples are similar to those for the whole sample of clusters (top panel of Table 4). In particular, we recover in both subsamples the order of the intrinsic scatters ( $\sigma_g < \sigma_* < \sigma_{\text{BCG}}$ ) and positive correlation coefficients  $r_{*,\text{BCG}}$  and  $r_{l,g}$ . The intrinsic scatter of the total and BCG stellar components in the C2 sample is larger than



**Fig. 7.** The gas (green), total stellar (red), and BCG (magenta) mass fractions as a function of halo mass. The solid, dashed, and dotted lines denote the Bayesian estimation for the C1 clusters, the C2 clusters, and the whole sample, respectively. The filled and hatched regions denote the  $1\sigma$  uncertainties for the 83 C1 clusters and the 53 C2 clusters, respectively.

that in the C1 sample. The differences of  $\sigma_*$  and  $\sigma_{\text{BCG}}$  are at  $3.3\sigma$  and  $4.1\sigma$  levels, respectively.

### 3.7 Subsample with central radio sources

We split the sample into clusters with and without central radio sources associated with active galactic nucleus (AGN). We search for central radio sources within 60 arcsec from the BCGs using the FIRST (White et al. 1997) and TGSS (Intema et al. 2017) surveys. We find central radio sources in 34 clusters. The fraction of central radio sources is 0.25 and almost constant over the redshift. Average fractions to include radio sources within 60 arcsec from random positions and random galaxies of which  $z$ -band magnitudes are brighter than 20 ABmag are only 0.04 and 0.09, respectively. It is difficult to identify whether radio sources are associated with cluster members or not because of their extended distribution and a lack of their redshifts. A visual inspection of radio sources and optical distribution suggests that a contamination of radio overlapped at different redshifts is small. Even when we exclude two clusters whose central radio sources have a possibility to be overlapped with point sources at  $z \sim 1 - 1.5$ , we find consistent results.

We repeat the Bayesian analysis for the two subsamples with and without central AGNs. We refer to the former and latter samples as radio-AGN (R) clusters and non-radio-AGN (NR) clusters, respectively. The resulting regression parameters for the 34 radio-AGN clusters are similar to those of the 102 non-radio-AGN clusters (bottom panel Table 3). Since the errors are large, it is difficult to discriminate between the two subsamples, as shown in Figure 8. The intrinsic covariance between the baryon components for the two subsamples are similar to those for the whole sample of clusters (bottom panel of Table 4).

**Table 3.** The top and bottom panels show the best-fit scaling relations for the 83 C1 clusters and the 53 C2 clusters and the 102 non-radio-AGN (NR) clusters and the 34 radio-AGN (R) clusters, respectively. The normalization,  $\alpha$ , and the slope,  $\beta$ , are defined by the linear regressions (eqs. 5 and 6).  $\dagger$ : the results using a trivariate Gaussian prior as the WL mass calibration, as described in Sec. 2.4.3. NR and R expresses "non-radio-AGN" and "radio-AGN", respectively. The errors denote the  $1\sigma$  uncertainty.

	$\alpha$	$\beta$
C1 : $M_{500}^{\text{WL}} E(z)$	$-0.11^{+0.02\dagger}_{-0.02}$	$1.09^{+0.02\dagger}_{-0.02}$
$L_X E(z)^{-1}$	$0.30^{+0.18}_{-0.18}$	$1.73^{+0.34}_{-0.47}$
$M_* E(z)$	$0.78^{+0.11}_{-0.09}$	$0.89^{+0.14}_{-0.13}$
$M_{\text{BCG}} E(z)$	$-0.93^{+0.09}_{-0.08}$	$0.53^{+0.13}_{-0.13}$
$M_g E(z)$	$1.97^{+0.10}_{-0.10}$	$1.48^{+0.20}_{-0.25}$
C2 : $M_{500}^{\text{WL}} E(z)$	$-0.11^{+0.02\dagger}_{-0.02}$	$1.08^{+0.02\dagger}_{-0.02}$
$L_X E(z)^{-1}$	$0.11^{+0.28}_{-0.23}$	$1.65^{+0.34}_{-0.23}$
$M_* E(z)$	$0.60^{+0.17}_{-0.15}$	$0.92^{+0.19}_{-0.14}$
$M_{\text{BCG}} E(z)$	$-1.02^{+0.17}_{-0.16}$	$0.53^{+0.18}_{-0.15}$
$M_g E(z)$	$1.87^{+0.16}_{-0.14}$	$1.50^{+0.23}_{-0.15}$
NR : $M_{500}^{\text{WL}} E(z)$	$-0.11^{+0.02\dagger}_{-0.02}$	$1.09^{+0.02\dagger}_{-0.02}$
$L_X E(z)^{-1}$	$0.32^{+0.18}_{-0.16}$	$1.48^{+0.32}_{-0.23}$
$M_* E(z)$	$0.75^{+0.10}_{-0.09}$	$0.91^{+0.13}_{-0.10}$
$M_{\text{BCG}} E(z)$	$-0.92^{+0.10}_{-0.10}$	$0.54^{+0.14}_{-0.12}$
$M_g E(z)$	$1.96^{+0.11}_{-0.10}$	$1.37^{+0.20}_{-0.14}$
R : $M_{500}^{\text{WL}} E(z)$	$-0.11^{+0.02\dagger}_{-0.02}$	$1.09^{+0.02\dagger}_{-0.02}$
$L_X E(z)^{-1}$	$0.03^{+0.26}_{-0.31}$	$1.71^{+0.43}_{-0.39}$
$M_* E(z)$	$0.72^{+0.14}_{-0.15}$	$0.87^{+0.25}_{-0.20}$
$M_{\text{BCG}} E(z)$	$-0.92^{+0.11}_{-0.12}$	$0.38^{+0.19}_{-0.17}$
$M_g E(z)$	$1.88^{+0.14}_{-0.17}$	$1.42^{+0.26}_{-0.21}$

### 3.8 Redshift evolution

We next investigate the redshift evolution of the baryon budget. We here define the observables independent of redshifts, as follows,

$$x = \ln \frac{M_{500}^{\text{WL}}}{10^{14} M_\odot} \quad (10)$$

$$\mathbf{y} = \left\{ \ln \frac{L_X}{10^{43} \text{ erg s}^{-1}}, \ln \frac{M_*}{10^{12} M_\odot}, \ln \frac{M_{\text{BCG}}}{10^{12} M_\odot}, \ln \frac{M_g}{10^{12} M_\odot} \right\}.$$

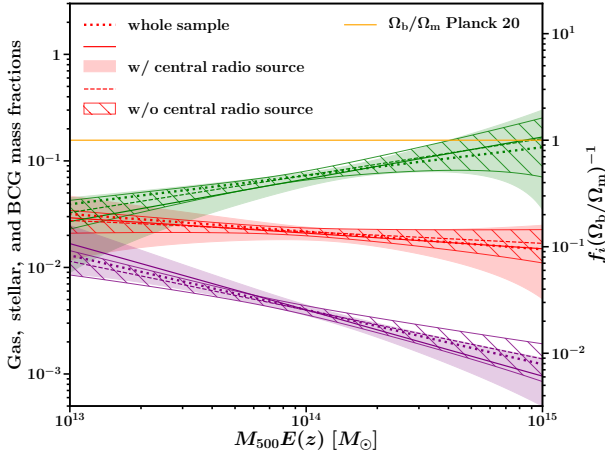
We assume the following redshift dependence of the scaling relations,

$$\mathbf{Y}_Z = \boldsymbol{\alpha} + \boldsymbol{\beta} Z + \boldsymbol{\gamma} \ln \left( \frac{E(z)}{E(0.3)} \right), \quad (11)$$

with  $Z = \log(M_{500}/10^{14} M_\odot)$ . We repeat the Bayesian analysis for the multivariate scaling relations. We assume that the redshift dependence of the  $L_X - M$  relation follows a self-similar

**Table 4.** The top and bottom panels show the intrinsic covariance for the 83 C1 clusters and the 53 C2 clusters and the 102 non-radio-AGN (NR) clusters and the 34 radio-AGN (R) clusters, respectively. Each column is the same as in Table 2.

	$L_X E(z)^{-1}$	$M_* E(z)$	$M_{\text{BCG}} E(z)$	$M_g E(z)$
C1 : $L_X E(z)^{-1}$	$0.38^{+0.33}_{-0.17}$	$-0.02^{+0.08}_{-0.03}$	$0.01^{+0.06}_{-0.04}$	$0.06^{+0.18}_{-0.04}$
$M_* E(z)$	$-0.15^{+0.47}_{-0.39}$	$0.36^{+0.09}_{-0.06}$	$0.06^{+0.04}_{-0.03}$	$-0.01^{+0.07}_{-0.03}$
$M_{\text{BCG}} E(z)$	$0.04^{+0.22}_{-0.29}$	$0.31^{+0.14}_{-0.17}$	$0.54^{+0.05}_{-0.05}$	$-0.01^{+0.04}_{-0.03}$
$M_g E(z)$	$> 0.82$	$-0.12^{+0.61}_{-0.47}$	$-0.13^{+0.30}_{-0.34}$	$< 0.38$
C2 : $L_X E(z)^{-1}$	$0.25^{+0.13}_{-0.08}$	$-0.05^{+0.06}_{-0.06}$	$-0.04^{+0.08}_{-0.08}$	$0.03^{+0.03}_{-0.02}$
$M_* E(z)$	$-0.31^{+0.40}_{-0.33}$	$0.66^{+0.08}_{-0.09}$	$0.44^{+0.11}_{-0.11}$	$-0.01^{+0.06}_{-0.05}$
$M_{\text{BCG}} E(z)$	$-0.22^{+0.38}_{-0.34}$	$0.78^{+0.06}_{-0.09}$	$0.87^{+0.07}_{-0.08}$	$-0.01^{+0.07}_{-0.07}$
$M_g E(z)$	$0.71^{+0.17}_{-0.43}$	$-0.06^{+0.50}_{-0.43}$	$-0.06^{+0.48}_{-0.44}$	$< 0.21$
NR : $L_X E(z)^{-1}$	$0.70^{+0.17}_{-0.27}$	$0.02^{+0.15}_{-0.09}$	$0.07^{+0.12}_{-0.10}$	$0.25^{+0.17}_{-0.16}$
$M_* E(z)$	$0.07^{+0.32}_{-0.35}$	$0.49^{+0.10}_{-0.07}$	$0.25^{+0.08}_{-0.07}$	$0.02^{+0.11}_{-0.06}$
$M_{\text{BCG}} E(z)$	$0.16^{+0.17}_{-0.22}$	$0.68^{+0.07}_{-0.13}$	$0.74^{+0.07}_{-0.06}$	$0.03^{+0.08}_{-0.06}$
$M_g E(z)$	$> 0.95$	$0.11^{+0.40}_{-0.44}$	$0.12^{+0.20}_{-0.26}$	$0.38^{+0.11}_{-0.17}$
R : $L_X E(z)^{-1}$	$0.31^{+0.21}_{-0.12}$	$-0.01^{+0.07}_{-0.06}$	$-0.02^{+0.07}_{-0.06}$	$0.04^{+0.05}_{-0.02}$
$M_* E(z)$	$-0.04^{+0.37}_{-0.43}$	$0.50^{+0.11}_{-0.09}$	$0.13^{+0.08}_{-0.06}$	$-0.00^{+0.05}_{-0.04}$
$M_{\text{BCG}} E(z)$	$-0.13^{+0.37}_{-0.37}$	$0.53^{+0.13}_{-0.20}$	$0.53^{+0.09}_{-0.07}$	$-0.03^{+0.04}_{-0.03}$
$M_g E(z)$	$0.83^{+0.11}_{-0.33}$	$-0.02^{+0.49}_{-0.43}$	$-0.27^{+0.42}_{-0.34}$	$0.17^{+0.09}_{-0.03}$



**Fig. 8.** The gas (green), total stellar (red), and BCG (magenta) mass fractions as a function of halo mass. The solid, dashed, and dotted lines denote the Bayesian estimation for the 34 radio-AGN clusters, the 102 non-radio-AGN clusters, and the whole sample, respectively. The filled and hatched regions denote the  $1\sigma$  uncertainties for the 34 radio-AGN clusters and the 102 non-radio-AGN clusters, respectively.

solution with  $\gamma_l = 2$  in eq. 11 to infer the redshift-dependent parent population  $p(Z, z)$ . Table 5 summarizes the resulting regression parameters. The resulting normalization and mass-dependent slopes are in good agreement with those for the self-similar redshift evolution (Table 1). We find no redshift evolution in the  $M_* - M$ ,  $M_{\text{BCG}} - M$ , and  $M_g - M$  relations which agree well with the self-similar redshift evolution  $\gamma = 0$ .

## 4 Discussion

### 4.1 Systematics

We recall the method of the multivariate-scaling-relations analysis of the baryon components. We set the vectors of baryons in  $\mathbf{y}$ , weak-lensing mass in  $x$  (Sec. 2.4) and the error covariance matrix (Sec. 2.4.1). The weak-lensing mass and true mass are statistically related through a power-law relation with intrinsic scatter based on a prior motivated by numerical simulations (Sec. 2.4.3). Our Bayesian method (Sec. 2.4.2) simultaneously computes the linear regression parameters ( $\alpha$  and  $\beta$ ), the intrinsic covariance ( $\Sigma_{\text{int}}$ ), and the parent population of the true mass ( $p(Z, z)$ ). In the regression analysis, it is vitally important to control regression dilution effect and selection effect (see details in Appendix A). The two effects are simultaneously calibrated by the estimated parameters of the assumed parent population (Sec. 2.4.2 and Appendix A). The shape of the parent population depends on the weak-lensing mass distribution as well as

the X-ray luminosity which is used as an approximated tracer of our cluster finder. This subsection discusses possible sources of systematics in the Bayesian regression analysis.

#### 4.1.1 Performance of Bayesian analysis

We assess the reliability of our Bayesian analysis using mock simulations computed with the error matrix similar to the observational one (see details in Appendix A). We define a multiplicative error and an additive error in the relation  $\theta_{\text{output}} = m \times \theta_{\text{input}} + c$ , where  $\theta_{\text{input}}$  and  $\theta_{\text{output}}$  are the input and the output parameters. The resulting multiplicative and additive errors in the simulation of  $p(Z, z)$  are  $\langle m \rangle = 0.989 \pm 0.018$  and  $\langle c \rangle = -0.004 \pm 0.011$  averaged over all the parameters, indicating that our code recovers well the input parameters. The uncertainties for the estimated parameters of the multivariate scaling relations are larger than the accuracy of the recovery of the input parameters. In the case of the large measurement errors of the weak-lensing masses, it is important to carefully and fully consider the error correlations. If larger errors of two observables are not correlated, it is difficult to constrain the intrinsic covariance (Figure 14 in Appendix A). The stacked observables sorted by X-ray luminosities and inverse-weighted with  $\Sigma_{\text{err}} + \Sigma_{\text{int}}$  (Figure 2-4; Sec. 3.1) are in good agreement with the baselines described by  $\alpha$  and  $\beta$ , which ensures the consistency between the independent regression parameters to represent the data.

#### 4.1.2 Parent population

We employed the Gaussian distribution for the parent population,  $p(Z, z)$ , for the logarithm of the true mass. This functional formulae differs from the XXL X-ray luminosity function (Pacaud et al. 2016; Adami et al. 2018; Valotti et al. 2018). We infer the shape parameters ( $\mu_Z(z)$  and  $\sigma_Z(z)$ ) of  $p(Z, z)$  by the hierarchical Bayesian modelling, which keeps a flexibility to approximately describe an unknown parent population or a halo mass function of Tinker et al. (2008). It effectively corrects both the regression dilution effect and the selection bias (see Appendix A and B). Furthermore, our results are not affected

by cluster mass function and the tension in  $\sigma_8$  measurements between *Planck* early universe experiment and nearby universe observations (Pratt et al. 2019). We emphasize that the purpose to introduce  $p(Z, z)$  is not to constrain cosmological parameters or to accurately determine the mass function but to correct the above two effects in the analysis of the multivariate scaling relations (Appendix A and B).

The XXL selection function behind the XXL cluster catalog uses the actual surface brightness profile, namely, core radius and total count-rate, to avoid contamination by X-ray point sources (Pacaud et al. 2016; Adami et al. 2018; Valotti et al. 2018). The total X-ray luminosity is computed by integrating the X-ray surface brightness distribution. We can easily infer  $\sigma_Z(z)$  and  $\mu_Z(z)$  through the  $L_X - M$  relation, which is sufficient to constrain the regression parameters of the baryon contents for the current sample, as seen in Figures 2-4 and Appendix A. To accurately measure the  $p(Z, z)$  distribution, we could use multiple Gaussian distributions with different weights. However, when we used the double Gaussian distribution, we were not able to constrain the parameters of the second Gaussian distribution (Sec. 2.4.2). Therefore, the single Gaussian is sufficient to describe the multivariate scaling relations for the current sample. With a larger sample of clusters and/or small measurement errors of the weak-lensing masses, we would require a more sophisticated model such as the cluster mass function combined with the XXL selection function.

Since the X-ray selection is affected by cosmological dimming and the selected cluster masses depend on the redshift, we introduced the redshift-dependent mean  $\mu_Z(z)$  and standard deviation  $\sigma_Z(z)$  of  $p(Z, z)$  as hyper-parameters. Figure 9 shows the resulting X-ray luminosity population as a function of  $E(z)$  which is computed by  $p(Y_0|Z, z)p(Z, z)$ . The resulting models of the X-ray luminosity population estimated by the true mass distributions of  $p(Z, z)$  agree with the data distribution. The resulting models indicate that the XXL cluster catalog covers a wide range of X-ray luminosities at low redshifts and comprises only the most X-ray luminous clusters at high redshifts. The X-ray luminosity population for the C1 clusters is shifted to a higher value compared to that for the C2 clusters. The whole X-ray luminosity population is distributed around the intermediate position between the two C1 and C2 X-ray population.

As an alternative modelling, we here assume a redshift-independent Gaussian distribution  $p(Z) = \mathcal{N}(\mu_Z, \sigma_Z)$  in the Bayesian analysis, where the mean ( $\mu_Z$ ) and standard deviation ( $\sigma_Z$ ) are free parameters independent of the redshift. We refer it to as no- $z$ . We compare the models using Akaike's information criterion (AIC) and Bayesian information criterion (BIC). The AIC and BIC are defined by  $\text{AIC} = 2N_{\text{para}} - 2\ln p_{\text{max}}$  and  $\text{BIC} = N_{\text{para}} \ln N_{\text{data}} - 2\ln p_{\text{max}}$ , respectively. Here  $N_{\text{para}}$  is the number of parameters,  $N_{\text{data}}$  is the number of data points, and  $p_{\text{max}}$  is the maximum value of the posterior probability (eq.

**Table 5.** Regression parameters of the scaling relations with redshift evolution. The normalization,  $\alpha$ , the slope,  $\beta$ , and the redshift evolution,  $\gamma$ , as a function of the true mass are defined by the linear regression (eq. 11). The square bracket denotes the fixed value.  $\dagger$ : the results using a trivariate Gaussian prior as the WL mass calibration, as described in Sec. 2.4.3. The errors denote the  $1\sigma$  uncertainty.

	$\alpha$	$\beta$	$\gamma$
$M_{500}^{\text{WL}}$	$-0.11^{+0.02\dagger}_{-0.02}$	$1.08^{+0.02\dagger}_{-0.02}$	[0]
$L_X$	$0.33^{+0.14}_{-0.13}$	$1.28^{+0.22}_{-0.17}$	[2]
$M_*$	$0.75^{+0.13}_{-0.13}$	$0.80^{+0.11}_{-0.11}$	$-0.25^{+0.57}_{-0.49}$
$M_{\text{BCG}}$	$-1.03^{+0.15}_{-0.15}$	$0.41^{+0.11}_{-0.10}$	$-0.06^{+0.55}_{-0.54}$
$M_g$	$1.92^{+0.10}_{-0.10}$	$1.23^{+0.13}_{-0.10}$	$0.39^{+0.33}_{-0.32}$

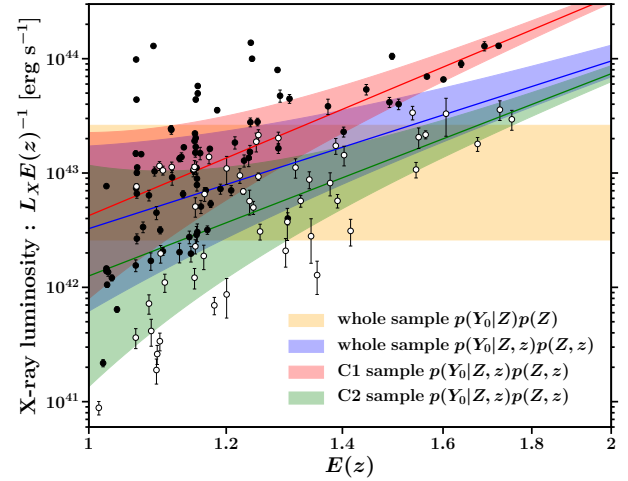
A2). The first terms in both the AIC and BIC describe a penalty of over-fitting by increasing the number of parameters in the model. The AIC, derived by relative entropy, measures relative loss among given different models. A low AIC value means that a model is considered to be closer to the truth. The BIC is derived by the framework of Bayesian theory to maximize the posterior probability of a model given the data. In other words, a model with the lowest BIC is preferred to be the truth. The AIC and BIC are based on different motivations and thus they provide complementary information. When we compare with two models, the model with the lower value is preferred. The difference between the models are significant according to both the AIC and BIC,  $\Delta\text{AIC} = \text{AIC}_{\text{no-}z} - \text{AIC} = +43$  and  $\Delta\text{BIC} = \text{BIC}_{\text{no-}z} - \text{BIC} = +35$ . The orange region in Figure 9 does not match with the data distribution. Therefore, the redshift-dependent parent population is preferable.

We next assume a second-order redshift dependence for  $\mu_Z(z) = \mu_{Z,0} + \gamma_{\mu_Z} \ln E(z) + \gamma_{\mu_Z,2} (\ln E(z))^2$  and  $\ln \sigma_Z(z) = \ln \sigma_{Z,0} + \gamma_{\sigma_Z} \ln E(z) + \gamma_{\sigma_Z,2} (\ln E(z))^2$ . The result does not significantly change and, thus, the resulting AIC and BIC become larger (worse) than those of our main result due to the penalty from the increased number of parameters;  $\Delta\text{AIC} = +3$  and  $\Delta\text{BIC} = +13$ . The first-order dependence is sufficient to describe the data.

The code could in principle estimate the  $\mu_Z(z)$  and  $\sigma_Z(z)$  parameters without a correction for the Malmquist bias. We perform the Bayesian analysis using a subset of  $\mathbf{y} = \{y_*, y_{\text{BCG}}, y_g\}$  in order to understand the impact of the tracer of cluster finders in the multivariate scaling relation analysis. The resulting  $(\sigma_Z(z))^2$  becomes higher by  $\sim 7\%$ , and consequently the mass-dependent slopes,  $\beta$ , become shallower by  $\sim 5\%$  in  $\beta_*$ ,  $\sim 7\%$  in  $\beta_{\text{BCG}}$ , and  $\sim 5\%$  in  $\beta_g$ . This change is caused by the relationship between the variance in the parent population and the slope (eq. A13), which is described in Appendix B. Changes in the normalization is less than 1%. Although the overall results do not change, the simultaneous treatment of the X-ray luminosity approximately related to the cluster finding can more properly estimate the parent population in the computation of the multivariate scaling relations.

#### 4.1.3 Systematics by weak-lensing mass calibration

In the limit of low S/N weak-lensing signals, the errors of the weak-lensing masses are mainly caused by the number of background galaxies, rather than intrinsic halo properties of the halo non-sphericity, subhalos and its surrounding large-scale structure. Indeed, Umetsu et al. (2020) have shown that the measurement errors using synthetic weak-lensing data of the analytic NFW model are comparable to those using cosmological simulations. We independently introduced the bias and scatter in weak-lensing mass measurement based on dark-matter-only simulations (Umetsu et al. 2020). Higher-mass halos tend



**Fig. 9.** The X-ray luminosity versus  $E(z)$ . Black and white circles are the C1 and C2 clusters, respectively. The color regions denote the standard deviation ranges of the X-ray luminosity population expected by the resulting parent population of true mass  $p(Z)$ . The solid lines are the mean of the X-ray luminosity population. Orange color is the redshift-independent case of  $p(Y_0|Z, z)p(Z, z)$ . Blue, red, and green colors are the whole, C1, and C2 samples with the redshift-dependent case of  $p(Y_0|Z, z)p(Z, z)$ .

to have less spherical structure because they are the most recent forming systems and thus growth of halos may not have yet erased the information about initial condition and formation process. (e.g. Jing & Suto 2002; Allgood et al. 2006). The abundance of subhalos in high-mass halos is larger than in low-mass halos. Therefore, the weak-lensing mass calibration inherent in the intrinsic halo properties depends on cluster masses (Umetsu et al. 2020).

Since it is difficult to calibrate weak-lensing masses for individual clusters in response to their own properties, we employed the statistical approach of the weak-lensing mass calibration (eq. 5). The overdensity radii are accordingly changed from the weak-lensing overdensity radii by the weak-lensing mass calibration. We adapted the weak-lensing overdensity radius as the measurement radius for the gas mass and the stellar mass because the individual true masses are not clarified. Since the weak-lensing overdensity radii are only  $-6 \pm 4$  percent smaller than those of the true mass, it is negligible compared to the statistical errors.

In Sec 3.3, we estimated how much the gas mass and the stellar mass are changed by the aperture correction which is caused by the difference in the measurement radii of  $r_{500}^{\text{WL}}$  and  $r_{500}$ . Since the mass bias becomes larger with decreasing the true mass (Fig. 1), the aperture correction makes the stellar mass and the gas mass 5% and 9% higher at  $M_{500}E(z) \sim 10^{13} M_\odot$ , respectively. In contrast, the aperture correction at the high mass end of  $\sim 10^{15}$  is less than a few percent.

When we fix the estimated values of  $\alpha_{\text{WL}}, \beta_{\text{WL}}, \sigma_{\text{WL}}$  of the weak-lensing mass calibration ignoring their uncertainties, the measurement errors of the gas mass scaling relation at

$M_{500}E(z) \sim 10^{15} M_{\odot}$  and the intrinsic correlation coefficient are reduced by 10 – 30 percent, while the other errors are not significantly changed. When we remove the prior of  $\sigma_{\text{WL}}$ , we obtain  $\sigma_{\text{WL}} = 0.14^{+0.09}_{-0.06}$  which is consistent with the mass calibration determined by the 639 simulated clusters.

We also investigate the multivariate scaling relation with the mass calibration with the tanh function (Sec. 2.4.3). In that case, we replace  $M_{500}^{\text{WL}}$  by the calibrated mass,  $M_{500}^{\text{calib}}$ , in the  $X$  quantity and treat  $X = Z$  with a fixed  $\sigma_X = \sigma_{\text{WL}}$ . The resulting regression parameters are consistent with those of our main result (Table 1) within  $1\sigma$  errors. When we scale the  $y_*$  and  $y_g$  observables by  $\delta y \simeq \delta x$  according to the mass calibration, the result does not change significantly.

We fix the intrinsic correlation coefficients of  $r_{\text{WL},i} = 0$  in the analysis. When we treat them as free parameters, the fitting results are almost the same. Therefore, an over-fitting by increasing the number of free parameters occurs. A penalty of the over-fitting increases the information criterion, especially  $\Delta\text{BIC} = +17$ . We thus prevent over-fitting by fixing the correlations in the analysis.

#### 4.1.4 Systematics by error correlation

Since the member galaxies are sparsely distributed, we adopted a single error correlation coefficient  $r_{\text{WL},*}^{\text{err}}$  computed over the whole sample. Intrinsic covariance might be affected by this treatment. We therefore assess how much the intrinsic covariance is changed by our choice of  $r_{\text{WL},*}^{\text{err}}$ . We first pick a uniform random number from a range of  $[r_{\text{WL},*}^{\text{err}} - 0.1, r_{\text{WL},*}^{\text{err}} + 0.1]$  for each cluster and we find that the resulting intrinsic covariance is consistent with our reference results (Table 2). Therefore, our treatment does not significantly impact the results. Next, we use  $r_{\text{WL},*}^{\text{err}} = -0.5, 0, 0.5$  which are lower than 0.873 (Sec. 2.4.1), and we find that all the results of  $r_{*,g}$  become negative or no correlation in contrast to our positive result (Table 2), and the uncertainties for  $r_{\text{BCG},*}$  and  $r_{*,g}$  become larger by 2.2 and 1.6 times, respectively. The other parameters are not significantly affected by the assumption. The change of  $r_{*,g}$  is caused by  $r_{*,g}^{\text{err}} = r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}}$ . Therefore, an improper treatment can give rise to spurious anti-correlation between gas mass and stellar mass.

#### 4.1.5 Blue galaxies and the intracluster light

We counted the total stellar masses of the red galaxies selected by the color-magnitude planes using the XXL centers and redshifts. In general, cluster members are composed of red and blue galaxies which are distributed in the inner and outer regions, respectively (e.g. Whitmore et al. 1993; De Propris et al. 2004; Nishizawa et al. 2018). Red galaxies would be the dominant component of the cold baryon within  $r_{500}$  which is roughly about half of virial radius. As for blue galaxies distributed at outer radii, there is the possibility of an over-subtraction of

background component. We estimate how much stellar mass is changed when including blue galaxies. We first select galaxies from the MIZUKI photometric redshift catalog (Nishizawa et al. 2020; Aihara et al. 2019; Tanaka et al. 2018; Tanaka 2015) with criterion of  $|z_{\text{ph}} - z_c| < 0.05(1 + z_c)$  and  $M_* > 10^{10} M_{\odot}$ , where  $z_{\text{ph}}$  is a photometric redshift. We pick-up galaxies which are not identified in the red galaxy catalog but in the photometric catalog and refer them to as blue galaxies. When we include blue galaxies, the total stellar masses for individual clusters are changed by  $15 \pm 17$  percent. We repeat the Bayesian analysis for the red and blue galaxies and obtain  $\alpha_* = 0.70^{+0.09}_{-0.08}$ ,  $\beta_* = 0.81^{+0.13}_{-0.10}$ , and  $\sigma_* = 0.49^{+0.10}_{-0.06}$ . The baseline for the red and blue galaxies agrees with that for the red galaxies within the uncertainties (Table 1). The intrinsic correlation coefficients ( $r_{*,\text{BCG}} = 0.57^{+0.08}_{-0.10}$  and  $r_{*,g} = 0.29^{+0.34}_{-0.26}$ ) are not significantly changed, either.

The tidal stripping of stars from interacting galaxies and the merger of small galaxies with central brightest cluster galaxies make a diffuse intracluster light (ICL). In particular, extended low-surface brightness envelope forms around the central galaxies. However, it is very difficult to observationally detect such a weak excess of the ICL component from the image background because of over-subtraction. The HSC-SSP data is currently not adequate for the study of the ICL. Since we adopted the `cmodel` magnitude, we did not include the ICL component. Huang et al. (2018a) have studied how much the `cmodel` photometry underestimates stellar masses for massive galaxies at  $z < 0.5$ . They evaluated a difference between stellar masses estimated by the `cmodel` magnitude and surface mass density profiles out to 100 kpc without imaging stacking. The stellar mass within 100 kpc corresponds to the total stellar mass because 100 kpc aperture covers 5 – 10 times of effective radii. Huang et al. (2018a) found that a median  $M_*$  with the `cmodel` magnitudes underestimate the stellar masses only for massive galaxies ( $M_* > 10^{11.6} M_{\odot} \sim 4 \times 10^{11} M_{\odot}$ ) by  $\sim 0.1 - 0.15$  dex. Based on their results, we expect that the total stellar+ICL masses within 100 kpc aperture around the massive galaxies like BCGs would be  $\sim 1.3 - 1.4$  times higher than our `cmodel` estimates.

## 4.2 Baryon budget

This subsection is focused on the discussion of our measurements of baryon budgets of the clusters.

### 4.2.1 Baryon fractions

We found that the gas and stellar mass fractions increase and decrease with increasing halo mass (Figure 5 and Sections 3.1 and 3.3), respectively. This trend can be explained by a halo mass dependence of the star formation efficiency. The star formation efficiency in low-mass clusters and groups is expected to

be higher than that in high-mass clusters. In addition, tidal interactions among galaxies and the removal of the gas reservoir of galaxies by ram-pressure are more inefficient in low-mass clusters than in high-mass ones. Therefore, a larger fraction of the gas in low-mass clusters is consumed to form stars through cooling, while galaxy formation tends to be inhibited in high-mass clusters.

AGN feedback is also important to determine the baryon budget, because it heats the surrounding gas and suppresses star formation and more or less modifies radial distribution of gas, especially in low-mass clusters. Some AGNs especially in central galaxies are energetic enough to expel the gas material of stars out from the relatively shallow potential well of the low-mass clusters. The expelled gas in low-mass clusters is difficult to be re-accreted. Since the star formation activity does not change the total baryon fraction because of the mass conservation (e.g. Kravtsov et al. 2005), the total baryon fraction without AGN feedback is expected to be constant against the halo mass. However, the gas redistribution by AGN feedback could change halo mass dependence on the baryon and gas mass fractions. Therefore, the degree of balance between star formation and all the AGN activities throughout the entire cluster history controls the mass dependence of the baryon contents. The result that the total baryon fraction reaches to the cosmic mean baryon fraction  $\Omega_b/\Omega_m$  at high-mass halos of  $10^{15} M_\odot$  indicates that the high-mass halos are close to a closed-box in which the total baryon is confined. On the other hand, since  $f_b \sim 0.5(\Omega_b/\Omega_m)$  at low-mass clusters of  $\sim 10^{13} M_\odot$ , the low-mass halos are likely to be an open-box in which baryons are not conserved. This is likely caused by AGN feedback.

We investigated the baryon fractions for the clusters which currently host radio AGN activity and those for the other clusters (Figure 8 and Sec. 3.7). We do not find a significant difference of the baryon fraction in response to the current AGN activity. In general, AGN activity is a transient phenomenon, whereas gas ejection from the potential well depends on the total integrated non-gravitational energy. It implies that the cumulative quantities such as the gas and stellar masses are insensitive to the current AGN activity. However, a larger sample is essential to further constrain the parameters.

We also found good agreement of the baryon fractions between the C1 and C2 clusters (Figure 7), though the likelihood function of the XXL selection for the C1 class is different from that of the C2 class. This is promising for XXL X-ray cluster counts analyses of cosmological parameters.

#### 4.2.2 Intrinsic covariance in baryon content

Clusters move around the baselines in the scaling relations due to mass accretion, mergers, cooling, and AGN feedback. Since the baryonic evolution is an order of sound-crossing time, their positions in the scaling-relations instantly change. Their statis-

tical properties are observed as intrinsic covariance. If all the baryons were confined within the halo (closed-box), the intrinsic correlation coefficient between the gas mass and the stellar mass is expected to be negative because of  $\delta f_b = 0 = \delta f_g + \delta f_*$ . As we mentioned above, the anti-correlation appears only for the case of the improper treatment of the error covariance matrix. We found no evidence that the intrinsic correlation coefficient between gas mass and stellar mass in the whole sample is correlated or anti-correlated. It is generally very difficult to accept the null hypothesis that the true correlation is zero under a finite uncertainty. With the  $N$  sample, the constraint has to satisfy with  $|r_{*,g}| \lesssim 0.168(N/136)^{-1/2}$  so that the  $p$ -value of the null hypothesis can be higher than 5 percent. The required uncertainty is about half of the current constraint. However, the margin of the error includes the null correlation. Our result does not contradict with the open-box scenario suggested by the total baryon fraction (Sec 4.2.1).

The  $M_{\text{BCG}} - M_{500}$  relation showed the largest intrinsic scatter and a weak-mass dependence, which implies a presence of another factor besides the halo mass in the BCG mass growth. We found significant correlations between  $M_*$  and  $M_{\text{BCG}}$  and between  $M_{\text{sat}}$  and  $M_{\text{BCG}}$  (Sec 3.5), qualitatively suggesting that the BCGs co-evolve with the satellite galaxies.

Intrinsic correlation between soft-band  $L_X$  and  $M_g$  is close to unity which is naturally explained by the X-ray emissivity.

The C2 clusters have larger intrinsic scatter of total and BCG stellar masses than the C1 clusters. The discrepancies for the former and latter cases are  $3.3\sigma$  and  $4.1\sigma$ , respectively. Since the C2 class has lower X-ray luminosity (lower masses), the stellar properties in low X-ray clusters are likely to be more diverse from cluster-to-cluster.

### 4.3 Comparison of numerical simulations

Recent cosmological hydrodynamic simulations (e.g. Young et al. 2011; McCarthy et al. 2011; Planelles et al. 2013; Martizzi et al. 2014; Le Brun et al. 2014; Wu et al. 2015; Sembolini et al. 2016a; McCarthy et al. 2017; Barnes et al. 2017b; Farahi et al. 2018a; Henden et al. 2020; Farahi et al. 2020) studied stellar mass and gas distributions in clusters and/or groups. The simulations include the effect of cooling, AGN feedback, star formation, and SN feedback and compare them with the results of non-radiative simulations. Since gas distributions are radially modified by AGN feedback, the scaling relations depends on overdensity radius (e.g. Young et al. 2011; Farahi et al. 2018a). Thus, when we compare numerical simulation with observations, it is important to choose the same overdensity as observations (i.e.  $\Delta = 500$ ). Results of numerical simulations depend on the different AGN models (e.g. McCarthy et al. 2011; Le Brun et al. 2014; Sembolini et al. 2016b).

### 4.3.1 Scaling relations and mass-dependent slopes

We compare our results with some of the simulations in Figures 10. Simulations results are rescaled to  $z = 0.3$ , close to the median redshift of the XXL clusters, assuming self similar evolution. We find that our normalization and slope in  $M_g - M$  and  $M_* - M$  relations broadly agree with those of numerical simulations over two orders of magnitude in mass.

We compare the slopes in the scaling relations at  $\Delta = 500$  (Figure 11). The slopes of some numerical simulations depend on the halo mass. For instance, Farahi et al. (2020) showed that the slopes of the gas mass and stellar mass scaling relations at massive clusters ( $M_{500} \sim 10^{15} M_\odot$ ) are close to unity and becomes steeper and shallower with decreasing mass, respectively. We therefore estimate the average value and scatter with a weight of the resulting parent population  $p(Z, z)$  to fairly compare with their values in our mass range. We use the results being as close as possible to our median redshift and consider the redshift dependence of  $p(Z, z)$ . The gas mass slopes of numerical simulations (Young et al. 2011; Planelles et al. 2013; Barnes et al. 2017b; Wu et al. 2015; Truong et al. 2018; Farahi et al. 2018a; Henden et al. 2020; Farahi et al. 2020) are higher than predicted by the self-similar model ( $\beta = 1$ ). Some simulations are slightly steeper than the self-similar expectation ( $1 < \beta_g < 1.1$ ) while others have a clear higher slope ( $\beta_g \gtrsim 1.2$ ). The simulation results are not converged. Our results agree with the former case (e.g. Young et al. 2011; Barnes et al. 2017a; Farahi et al. 2018a). The slopes of the stellar to total mass relation (Wu et al. 2015; Farahi et al. 2018a; Henden et al. 2020; Farahi et al. 2020; Pillepich et al. 2018) are less than unity and agree with our results. The steep gas slope and the shallow stellar slope are consistent with the physical interpretation that the star formation efficiency is higher in low-mass systems than in high-mass ones (Sec 4.2.1).

The  $M_{\text{BCG}} - M$  relation was studied in several numerical simulations (e.g. Le Brun et al. 2014; Cooper et al. 2015; Pillepich et al. 2018; Farahi et al. 2020; Henden et al. 2020). The evolution of BCGs depends on both star formation efficiency and AGN feedback. In addition, since the BCG is located near cluster center, galaxy-galaxy mergers are an important process for its fast growth. Using the UniverseMachine simulation (Behroozi et al. 2019), Bradshaw et al. (2020) showed the shallow slope and large intrinsic scatter in the  $M_{\text{BCG}} - M_{200}$  relation. It is explained by that  $M_{\text{BCG}}$  is a function of not only halo mass but also halo formation time.

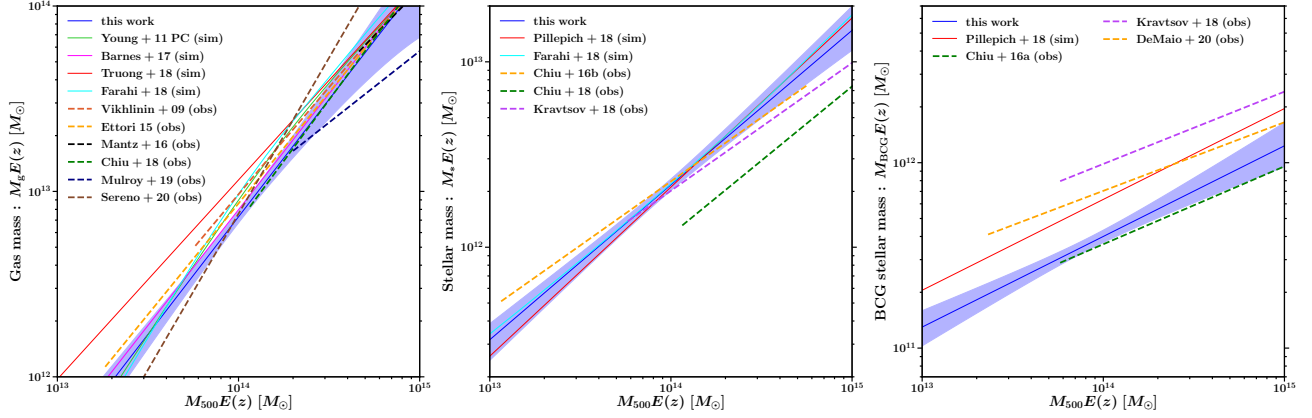
We compare the BCG-total mass slope with numerical simulations at  $\Delta = 500$ . The BCG mass slopes (Henden et al. 2020; Farahi et al. 2020; Pillepich et al. 2018) are much shallower than those of the  $M_* - M$  and  $M_g - M$  relations. We use stellar mass within three dimensional aperture of 30kpc (Pillepich et al. 2018) for a comparison. Although the BCG slopes show a diversity, they are similar to ours (Figure 11). We compare

with Pillepich et al. (2018) because the other two simulations do not show the normalization. They showed that the normalization of the BCG mass depends on the three-dimensional aperture size. The normalization with 100 kpc radius is about twice larger than that with 30 kpc. This is caused by the ICL component at the BCG outskirts. They also mentioned that the ICL stellar mass outside 30 kpc accounts for  $\sim 40$  percent of the total stellar mass of central galaxies and their surrounding ICL at  $M_{200} \sim 10^{13} M_\odot$  and  $\sim 80$  percent at  $M_{200} \sim 10^{15} M_\odot$ . We here compare with the normalization measured with 30kpc radius which is the minimum radius discussed in Pillepich et al. (2018) and covers the measurement regions of the `cmodel` magnitude. The normalization of the BCG mass to total stellar mass ratio of the numerical simulations (Figure 6; see also Figure 10) is constantly offset from our baseline by  $\sim 1.4$  times. We found that our BCG stellar mass estimates agree with those estimated by the CFHT photometry (Lavoie et al. 2016), as shown in Sec. 4.4.2. When we multiply the BCG mass by 1.3 because of the underestimation of the `cmodel` magnitude (Huang et al. 2018a, and Sec. 4.1.5), the discrepancy is improved. However, if we accordingly change the aperture size to 100 kpc, a factor 2-3 discrepancy between the observations and the simulation still remains. We leave for future work to understand the normalization offset.

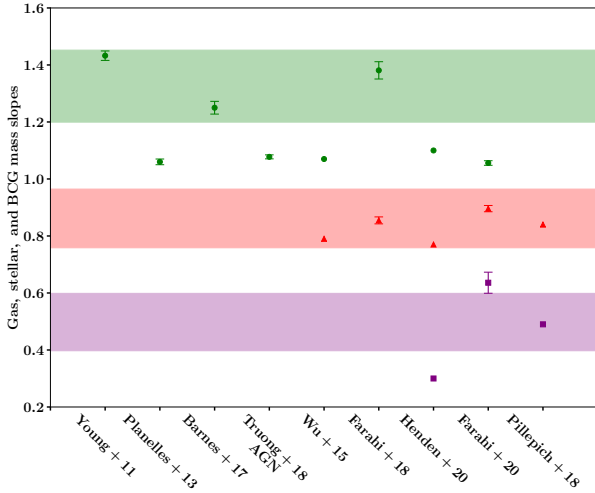
### 4.3.2 Intrinsic covariance in scaling relations

Intrinsic covariance is one of benchmarks to understand cluster evolution. Farahi et al. (2018a) computed how slope, normalization and intrinsic scatter change by a halo mass, using both BAHAMAS (McCarthy et al. 2017) and MACSIS (Barnes et al. 2017a) simulations. We weight their intrinsic scatters by  $p(Z, z)$  to derive representative values for our sample and obtain  $\langle \sigma_g \rangle = 0.20 \pm 0.02$  and  $\langle \sigma_* \rangle = 0.24 \pm 0.02$  at  $z = 0.5$ , where the errors are the  $1\sigma$  range over our mass range. Farahi et al. (2020) obtained the intrinsic scatter of  $\sigma_g = 0.065 \pm 0.003$ ,  $\sigma_* = 0.098 \pm 0.004$ , and  $\sigma_{\text{BCG}} = 0.333 \pm 0.015$  using IllustrisTNG simulations (Pillepich et al. 2018). There is a discrepancy between different numerical simulations. Although their scatter is somewhat lower than our results, the ascending order of intrinsic scatter of each baryon component is the same as our results:  $\sigma_g < \sigma_* < \sigma_{\text{BCG}}$ .

Some numerical simulations (e.g. Wu et al. 2015; Farahi et al. 2018a, 2020) showed that intrinsic correlation coefficient  $r_{*,g}$  at a fixed total mass is negative. Wu et al. (2015) showed a strong negative rank correlation  $r_{*,g} = -0.69$  between the deviations of the gas and stellar mass fractions from their baselines, although their definition is different from ours. The negative correlation appears in a wide overdensity range  $\Delta = 2500 - 10$ . They proposed a closed-box scenario where the intrinsic correlation coefficient  $r_{*,g}$  is anti-correlated. Farahi et al. (2018a) also found that intrinsic correlation coefficient changes with the



**Fig. 10.** Comparison of numerical simulations and observations in literature. From left to right, the gas, stellar, and BCG mass scaling relations. The solid blue line and regions denote the estimate and  $1\sigma$  uncertainty, respectively. Our results and numerical simulations denote the true masses in the all panels, while the masses for the previous observations except for Sereno et al. (2020) are not (mass proxies, lensing masses or hydrostatic masses). In the left panel, the solid green, magenta, red and light-blue lines are simulations (Young et al. 2011; Barnes et al. 2017a; Truong et al. 2018; Farahi et al. 2018b), and the dashed orange, yellow, black, green, dark blue, and brown lines are previous observations (Vikhlinin et al. 2009; Ettori 2015; Mantz et al. 2016; Chiu et al. 2018; Mulroy et al. 2019; Sereno et al. 2020), respectively. The line lengths of the observations denote their mass ranges. In the middle panel, the solid red and light-blue lines and the dashed yellow, green, magenta lines are simulations (Pillepich et al. 2018; Farahi et al. 2018b) and previous observations (Chiu et al. 2016b, 2018; Kravtsov et al. 2018), respectively. In the right panel, the solid red line and the dashed green, magenta, and yellow lines are a simulation (Pillepich et al. 2018) and observations (Chiu et al. 2016a; Kravtsov et al. 2018; DeMaio et al. 2020), respectively.



**Fig. 11.** Comparison of mass-dependent slopes  $\beta$  with those of numerical simulations (Young et al. 2011; Planelles et al. 2013; Barnes et al. 2017a; Truong et al. 2018; Wu et al. 2015; Farahi et al. 2018a; Henden et al. 2020; Farahi et al. 2020; Pillepich et al. 2018). Green, red, and magenta colour regions denote the  $1\sigma$  uncertainties of gas, stellar, and BCG mass slopes, respectively.

halo mass  $M_{500}$ . The intrinsic correlation at  $z = 0.5$  is nearly zero at  $M_{500} \sim 10^{13} M_{\odot}$ , and it is negative at  $\gtrsim 10^{14} M_{\odot}$ . They proposed that non-correlation at  $\sim 10^{13} M_{\odot}$  is caused by an open baryon box scenario in which the total baryon in low-mass clusters is not conserved by AGN feedback and proposed the closed-box scenario for the negative correlation for high-mass clusters. We recompute  $r_{*,g}$  at  $M_{500}$  from Figure 5 in Farahi et al. (2018a) with a weight of  $p(Z, z)$  and obtain  $\langle r_{*,g} \rangle = -0.21 \pm 0.05$ , where the second quantity is the

$1\sigma$  range. Farahi et al. (2020) also found a similar result  $-0.255 \pm 0.074$  measured at  $M_{200}$ . The probability of accidental correlations from 136 random pairs to realize the simulated result is  $\mathcal{O}(10^{-4})$ . Although we do not find such a negative correlation, a difference between their and our results is only  $1.4\sigma$  with our uncertainty.

Anbajagane et al. (2020) found a positive intrinsic correlation between central galaxy and total stellar masses at  $M_{200}$ . The intrinsic correlation coefficient weighted with  $p(Z, z)$ ,  $\langle r_{*,\text{BCG}} \rangle = 0.44 \pm 0.07$ ,  $0.44 \pm 0.04$ ,  $0.64 \pm 0.01$  and  $0.40 \pm 0.05$ , varies according to the simulation schemes. Farahi et al. (2020) found  $r_{*,\text{BCG}} = 0.273 \pm 0.052$  at  $M_{200}$ . Although the overdensity definitions are different, numerical simulations and our observation suggest that the mass growths of the total stellar mass and the BCG mass are correlated.

#### 4.3.3 Redshift evolution

Henden et al. (2020) investigated a redshift evolution in scaling relations. They found that the gas mass and the stellar mass at fixed halo mass increases and decreases with increasing redshift as  $M_g \propto (1+z)^{0.41 \pm 0.14}$  and  $M_* \propto (1+z)^{-0.51 \pm 0.08}$ , respectively, and the BCG mass weakly depends on the redshift  $M_{\text{BCG}} \propto (1+z)^{-0.15 \pm 0.10}$ . They concluded that the gas redshift evolution is attributed to the effectiveness of gas expulsion by AGN feedback with decreasing redshift. Le Brun et al. (2017) also found that the gas mass evolves with redshift as  $M_g \propto E(z)^{0.576 \pm 0.066}$ . On the other hand, Planelles et al. (2013) showed that redshift evolution for  $f_g$  and  $f_*$  are negligible. Truong et al. (2018) investigated a redshift evolution in X-ray scaling relations and they did not find a significant redshift

evolution in the  $M - M_g$  relation. Redshift evolution differs by different numerical simulations. Since our measurement errors of  $\gamma_g$  and  $\gamma_*$  are large, we cannot discriminate differences between numerical simulations.

#### 4.4 Comparison of observations

Gas and stellar masses in clusters were measured by various previous papers and projects (e.g. Lin et al. 2003, 2004; Okabe et al. 2010; Lin et al. 2012; Gonzalez et al. 2013; Laganá et al. 2013; Eckert et al. 2016; Zhang et al. 2016; Chiu et al. 2016b; Lin et al. 2017; Kravtsov et al. 2018; Chiu et al. 2018; Mulroy et al. 2019; Farahi et al. 2019; Sereno et al. 2020). Since each study adopted a different approach (Table 6), it is important to discuss differences in cluster sample, cluster mass measurement, and fitting procedure.

First, nowadays selection effects are more and more important. A cluster catalog can be constructed from optical (e.g. Rykoff et al. 2014; Oguri 2014; Rozo et al. 2016; Oguri et al. 2018; Maturi et al. 2019), X-ray (e.g. Böhringer et al. 2004; Piffaretti et al. 2011; Adami et al. 2018), thermal SZ effect (e.g. Planck Collaboration et al. 2014; Hilton et al. 2021; Bleem et al. 2021) or weak-lensing observations (e.g. Miyazaki et al. 2007, 2018a; Oguri et al. 2021). The Malmquist bias should be properly treated in fitting. Redshift ranges also vary with surveys. After all considerations, further differences could depend on intrinsic selection effects inherent in cluster astrophysics and/or different observational techniques.

Second, cluster mass measurements are one of important sources of systematic errors. To date, hydrostatic equilibrium mass, weak-lensing mass, or mass derived through scaling relations with mass proxies are used in the literature. A deviation from hydrostatic equilibrium, a lensing mass bias or intrinsic scatter in scaling relations (e.g. Pratt et al. 2019) should be considered in fitting methods (e.g. Sereno 2016).

Third, when the gas mass and stellar mass are measured within the same overdensity radii of the mass measurement, an error correlation between baryonic observables and mass should be considered in fitting (e.g. Okabe et al. 2010).

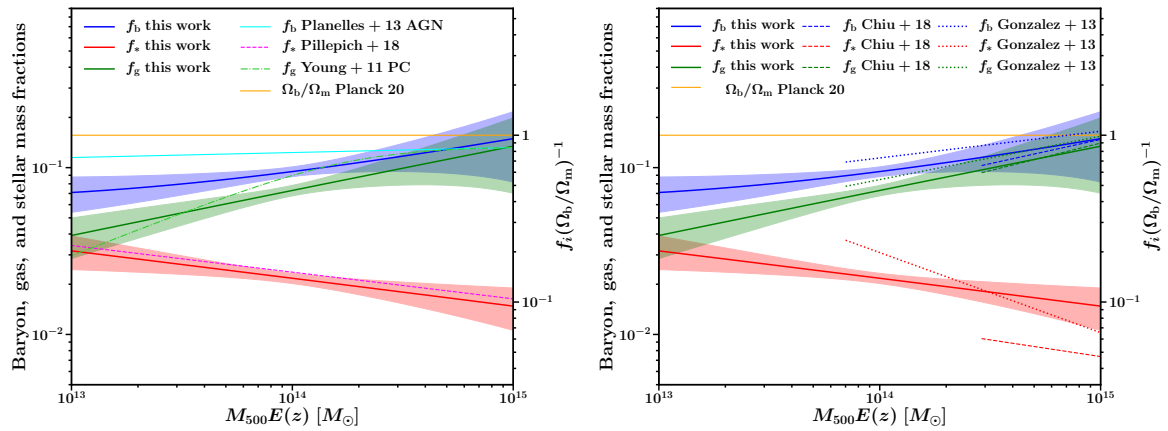
##### 4.4.1 Gas and stellar mass fractions

The left and middle panels of Figure 10 compare our results with the  $M_g - M$  and  $M_* - M$  relations from literature. Most of the previous papers analyzed several tens of clusters. Table 6 summarizes mass measurements and fitting method. In Figure 10, we use the Chabrier IMF for a comparison of stellar masses. The  $x$ -range of each line explicitly describes the mass range of each sample ( $M_{500} \gtrsim 10^{14} M_\odot$ ). We multiply the best-fit lines of the literature by  $E(z)$  when the literature uses  $M_g$  and  $M_{500}$  instead of  $M_g E(z)$  and  $M_{500} E(z)$ . Approaches of the previous papers can differ from ours. Nevertheless, the scaling rela-

tions broadly agree with our results. We stress the uniqueness of this study: the large sample of the 136 clusters with the nearly two orders of magnitude in mass including low-mass clusters of  $\mathcal{O}(10^{13} M_\odot)$  and our Bayesian analysis method fully considering the error covariance matrix, the selection effect, and the weak-lensing mass calibration.

We compare our  $M_g - M$  relation with the previous XXL papers (Eckert et al. 2016; Sereno et al. 2020). Eckert et al. (2016) studied the gas mass fraction for the 100 XXL brightest cluster sample (Pacaud et al. 2016). They used a temperature as mass proxy calibrated with a mass and temperature scaling relation based on a CFHTLenS (Heymans et al. 2012; Erben et al. 2013; Lieu et al. 2016). We have updated weak-lensing masses to the HSC-SSP shape catalog (Mandelbaum et al. 2018a; Umetsu et al. 2020). Since overdensity radii are also changed, we accordingly re-estimate gas mass and properly propagate the errors of weak-lensing mass in the scaling relation analysis. We also correct for the weak-lensing mass calibration (Umetsu et al. 2020). In the end, our main result of the  $M_g - M$  relation is about 30 percent higher than that of Eckert et al. (2016) at  $M_{500} \sim 10^{14} M_\odot$ . This point was already discussed in Sereno et al. (2020). Sereno et al. (2020) carried out the Bayesian analysis for the 118 XXL subset clusters using the HSC-SSP weak-lensing mass (Umetsu et al. 2020) and the gas mass measured within different radius. Their measurement radius is computed by an iterative procedure using the surface brightness profile and the  $f_g - M_{500}$  relation from Eckert et al. (2016). Their gas mass slope,  $1.55 \pm 0.30$ , is consistent with ours within errors, although their line in Figure 10 seems to be steeper. More conservatively, we repeat the regression analysis for the 118 clusters used in Sereno et al. (2020) and find  $\alpha_g = 1.96^{+0.07}_{-0.08}$  and  $\beta = 1.14^{+0.12}_{-0.10}$ . The difference in the slope is  $1.4\sigma$ , where  $\sigma$  is the error from Sereno et al. (2020). The normalization of Sereno et al. (2020) is  $\sim 4^{+1}_{-4}$  times higher than ours at  $M_{500} E(z) = 10^{15} M_\odot$  and  $\sim 0.4^{+0.6}_{-0.1}$  times of ours at  $10^{13} M_\odot$ . The gas mass fraction of Sereno et al. (2020) becomes higher than the cosmic baryon fraction  $\Omega_b/\Omega_m$  at  $M_{500} \gtrsim 4 \times 10^{14} M_\odot$ , while our result is lower at  $M_{500} \lesssim 10^{15} M_\odot$ . The two measurements are marginally consistent within the large error of Sereno et al. (2020).

We compare with gas mass scaling relations from Weighing the Giants (WtG; Mantz et al. 2016) and Local Cluster Substructure Survey (LoCuSS; Mulroy et al. 2019). The two surveys select their cluster samples from ROSAT All Sky Survey catalogues (Ebeling et al. 1998, 2000; Böhringer et al. 2004, RASS;) and comprise some of the most massive clusters ( $M_{500} \sim 4 \times 10^{14} M_\odot$ ). The redshift ranges for WtG and LoCuSS are redshifts  $0 < z < 0.5$  and  $0.15 < z < 0.3$ , respectively. They carried out analyses of multivariate scaling relations based on individual weak-lensing masses, taking into account both selection effect and cluster mass function. We can-



**Fig. 12.** Comparisons of numerical simulations (left) and previously observational studies (right). The solid blue, red and green lines are  $f_b$ ,  $f_*$  and  $f_g$  of this work, respectively. The orange horizontal line is  $\Omega_b/\Omega_m$  (Planck Collaboration et al. 2020). In the left panel, the solid light-blue, dashed magenta and long-dashed light-green lines are  $f_b$  (Planelles et al. 2013),  $f_*$  (Pillepich et al. 2018) and  $f_g$  (Young et al. 2011), respectively. In the right panel, the dashed blue, red, and green lines are  $f_b$ ,  $f_*$  and  $f_g$  for the SPT clusters (Chiu et al. 2018). The dotted blue, red, and green lines are  $f_b$ ,  $f_*$  and  $f_g$  for the Abell clusters (Gonzalez et al. 2013).

not find an explicit description about an aperture-induced error correlation between two baryonic observables (Sec 2.4.1; e.g.  $r_{*,g}^{\text{err}}$ ).

For WtG, the result of Mantz et al. (2016) is similar to ours. However, their slope using cluster data is  $\beta_g = 1.00 \pm 0.01$ . Their measurement error seems to be extremely small. Their slope is  $\sim 3\sigma$  lower than ours where  $\sigma$  is our measurement error.

For LoCuSS, Mulroy et al. (2019) gives shallower slope  $\beta_g = 0.7 \pm 0.1$ , which is a  $\sim 6\sigma$  difference from ours. When they applied the Bayesian code of Kelly (2007) to single  $M_g - M$  scaling relation, the slope  $\beta_g = 0.99^{+0.13}_{-0.14}$  becomes consistent with the slope of WtG. Therefore, the shallow gas mass slope of their main result might be related to their fitting code. When we adopt the result of Kelly (2007) for Mulroy et al. (2019), the gas mass slopes for massive clusters are close to unity. Numerical simulations (Farahi et al. 2018a, see also Sec 4.3.1) pointed out that the gas mass slope for very massive clusters  $\sim 10^{15} M_\odot$  is close to unity and becomes steeper with decreasing mass. The difference of slopes between ours and WtG/LoCuSS can be ascribed to the difference of halo mass. Since the two papers (Mantz et al. 2016; Mulroy et al. 2019) did not include stellar masses in their analyses, we cannot compare with them.

We next compare to the stellar mass fractions from literature. The XMM-Newton-Blanco Cosmology Survey (XMM-BCS; Chiu et al. 2016b) estimated the  $M_* - M_{500}$  relation for a sample of 46 clusters based on masses estimated through a scaling relation and found a shallow slope,  $M_* \propto M_{500}^{0.69 \pm 0.15} (1+z)^{-0.04 \pm 0.47}$ , with negligible redshift evolution. Lin et al. (2017) studied the  $M_* - M_{200}$  relation for optically selected clusters (CAMIRA; Oguri et al. 2018) using HSC-SSP data and found that the relation for the optical clusters are similar to those of X-ray selected clusters (Lin et al. 2012). Gonzalez

et al. (2013) studied gas mass and stellar mass fraction of optical Abell clusters (Abell et al. 1989) in the local Universe, with a hydrostatic-mass-based scaling relation. Their photometry includes ICL, but their stellar mass fraction is about half of ours. Chiu et al. (2018) studied baryon budget in 91 SZ clusters ( $0.2 < z < 1.25$ ) selected by the South Pole telescope (SPT) using  $M_{500}$  masses obtained through a scaling-relation and found  $M_* \propto M_{500}^{0.80 \pm 0.12} (1+z)^{0.05 \pm 0.27}$  (see also Figure 10). Their stellar mass slope agrees with ours, though the normalization of their stellar mass fraction is about 70 percent lower than ours. Decker et al. (2019) studied stellar mass fraction for massive ( $M_{500} \gtrsim 2 \times 10^{14} M_\odot$ ) and high-redshift ( $z \gtrsim 0.93$ ) clusters selected by an infrared survey, the Massive and Distant Clusters of the WISE Survey (MaDCoWS) and the SPT Survey. Their stellar mass fraction is consistent with optically-selected clusters (Gonzalez et al. 2013), and thus lower than ours. They also found that the difference of averaged stellar mass fraction between infrared and SZ selected sample is not significant.

The right panel of Figure 12 shows the gas and stellar mass fractions of the SZ (Chiu et al. 2018) and optically selected clusters (Gonzalez et al. 2013) for a comparison of different cluster finders (Table 6). Their mass ranges are higher than  $\sim 10^{14} M_\odot$ . The gas mass fraction in the SZ and optically selected clusters are similar to ours. We rescale the stellar mass fraction of Gonzalez et al. (2013) which used the Salpeter IMF by the Chabrier IMF. Although they include the component of the ICL, the stellar mass fraction for the optical clusters is lower than those in the XXL and SPT clusters. The gas mass fractions are similar to each other irrespective of the cluster finding methods, while the stellar mass fractions are slightly different.

**Table 6.** Summary of baryon fraction studies in literature and this study. <sup>a</sup>:cluster finders. X, SZ, opt, and mix represent X-ray, SZ, optical, and mixed (X/SZ/opt) clusters, respectively. Parentheses are samples defined by the XXL Survey, the ROSAT All-Sky Survey(RASS), XMM–Newton–Blanco Cosmology Survey (BCS), South Pole Telescope(SPT), or random sample (R), respectively. <sup>b</sup>:sample size. <sup>c</sup>:cluster mass measurement. WL, HE, and proxy represent weak-lensing mass, hydrostatic equilibrium mass, and mass estimated by scaling relations, respectively. <sup>d</sup>:typical mass in unit of  $M_\odot$ . The mass ranges are shown in Figures 10 and 12. <sup>e</sup> correction of mass bias in linear regression. <sup>f</sup>: scaling relation components (gas mass, total stellar mass, and BCG stellar mass). Some papers include other observables in multivariate scaling relations, which is denoted by <sup>†</sup>. <sup>g</sup>: consideration of selection effect in linear regression. <sup>h</sup>:consideration of the parent population. *G*, *K*, and mass represent Gaussian distribution, multiple Gaussian distributions using the Bayesian regression code of Kelly (2007) and mass function, respectively. <sup>i</sup>: aperture-induced error correlation between observable and measured mass. <sup>‡</sup> denotes a different approach. <sup>j</sup>:intrinsic covariance matrix for baryoninc components,  $\Sigma_{\text{int}}$ . full, diag, and partial denotes the full matrix, the diagonal terms, and the matrix including fixed elements.

	Clusters <sup>a</sup>	Size <sup>b</sup>	Mass <sup>c</sup>	Typical mass <sup>d</sup> [ $M_\odot$ ]	Mass bias <sup>e</sup>	Contents <sup>f</sup>	Selection effect <sup>g</sup>	$p(Z)$ <sup>h</sup>	Aperture $\Sigma_{\text{err}}$ <sup>i</sup>	$\Sigma_{\text{int}}$ <sup>j</sup>
<b>this work</b>	<b>X(XXL)</b>	<b>136</b>	<b>WL</b>	$\sim 10^{14}$	<b>yes</b>	$M_g/M_*/M_{\text{BCG}}$	<b>yes</b>	<b>G</b>	<b>yes</b>	<b>full</b>
Sereno et al. (2020)	X(XXL)	118	WL	$\sim 10^{14}$	yes	$M_g$ <sup>†</sup>	yes	<i>G</i>	-	full
Eckert et al. (2016)	X(XXL)	95	proxy	$\sim 10^{14}$	no	$M_g$	yes	<i>K</i>	no	diag
Mantz et al. (2016)	X(RASS)	27	WL	$\sim 8 \times 10^{14}$	yes	$M_g$ <sup>†</sup>	yes	mass	no <sup>‡</sup>	partial
Mulroy et al. (2019)	X(RASS)	41	WL	$\sim 6 \times 10^{14}$	yes	$M_g$ <sup>†</sup>	yes	mass	yes	full
Farahi et al. (2019)	X(RASS)	41	WL	$\sim 6 \times 10^{14}$	yes	$M_g$ <sup>†</sup>	yes	mass	yes	full
Vikhlinin et al. (2009)	X(RASS)	85	proxy	$\sim 3 \times 10^{14}$	no	$M_g$	yes	mass	no	diag
Chiu et al. (2016a)	SZ(SPT)	54	proxy	$\sim 6 \times 10^{14}$	no	$M_g/M_*/M_{\text{BCG}}$	no	no	no	diag
Chiu et al. (2018)	SZ(SPT)	91	proxy	$\sim 5 \times 10^{14}$	no	$M_g/M_*$	yes	mass	no	diag
Chiu et al. (2016b)	X(BCS)	46	proxy	$\sim 8 \times 10^{13}$	no	$M_*$	yes	mass	no	diag
Gonzalez et al. (2013)	opt(R)	15	proxy	$\sim 2 \times 10^{14}$	no	$M_g/M_*$	no	no	no	diag
Ettori (2015)	X(R)	59	HE	$\sim 3 \times 10^{14}$	no	$M_g$	no	no	yes	diag
Kravtsov et al. (2018)	opt(R)	9	proxy	$\sim 5 \times 10^{14}$	no	$M_*/M_{\text{BCG}}$	no	no	no	diag
Lavoie et al. (2016)	X(XXL)	85	proxy	$\sim 2 \times 10^{14}$	no	$M_{\text{BCG}}$	no	no	no	no
DeMaio et al. (2020)	mix(R)	23	proxy/WL	$\sim 10^{14}$	no	$M_{\text{BCG}}$	no	no	no	no

#### 4.4.2 $M_{\text{BCG}} - M$ relation

Literature results(Chiu et al. 2016a; Lavoie et al. 2016; Kravtsov et al. 2018; DeMaio et al. 2020) show that the mass-dependence slope in the  $M_{\text{BCG}} - M_{500}$  relation is  $\beta_* \sim 0.4 - 0.6$ . Analyses can differ by methodology: mass measurements, photometric measurements, inclusion or exclusion of the ICL. Nevertheless, the previous studies are comparable to our result within a factor 2-3 (Figure 10). Other papers (Lidman et al. 2012; Bellstedt et al. 2016; Zhang et al. 2016; Lin et al. 2017; Erfanianfar et al. 2019) using  $M_{200}$  also reported such a shallow slope.

Chiu et al. (2016a) measured the  $M_* - M$  relation without the ICL component. Both the slope,  $0.42 \pm 0.07$ , and the normalization are similar to ours.

DeMaio et al. (2020) measured the BCG stellar masses including the ICL component for 42 galaxy groups and clusters at  $z = 0.05 - 1.75$ , using *Hubble Space Telescope* (HST) data. The slope,  $0.37 \pm 0.05$ , is similar to ours, while the normalization is  $\sim 1.2 - 1.8$  times higher than ours at  $10^{15} M_\odot$  and  $2 \times 10^{13} M_\odot$ . They also found that the stellar envelope masses at  $10 \text{ kpc} < r < 100 \text{ kpc}$  are  $\sim 50 - 80\%$  of the stellar masses

within  $r < 100 \text{ kpc}$  at  $M_{500} \sim 2 \times 10^{13} - 10^{15} M_\odot$ . The normalization offset would be partially due to the ICL component.

Kravtsov et al. (2018) studied the  $M_{\text{BCG}} - M$  relation for nine nearby clusters at  $z \lesssim 0.1$ , using the SDSS DR8. They measured light profiles using raw images in order to account for the ICL component at the BCG outskirts and covert the stellar mass using the mass and light relation. The best-fit baseline for the nine clusters is shown by magenta dashed line in the right panel of Figure 10. Their slope,  $0.39 \pm 0.17$ , is similar to ours, while their normalization is about twice higher than ours. We compute the BCG masses for their clusters using the same data and find the weighted geometric mean ratio,  $\langle M_{\text{BCG}}^{\text{our}}/M_{\text{BCG}}^{\text{K}} \rangle = 0.42 \pm 0.01$ . When we use the SDSS DR16,  $\langle M_{\text{BCG}}^{\text{our}}/M_{\text{BCG}}^{\text{K}} \rangle = 0.53 \pm 0.02$ . The discrepancy between the normalization can be explained by a difference of treatment of the ICL component. Kravtsov et al. (2018) have also pointed out that their  $r$ -band luminosities are twice or more higher than those estimates by the *cmodel* magnitude, which is consistent with our comparison.

Lavoie et al. (2016) investigated the  $M_{\text{BCG}} - M$  relation for the XXL 100 brightest cluster sample (Pacaud et al. 2016) us-

ing the scaling-relation-based masses and found a steep slope  $\beta_{\text{BCG}} = 1/(0.84 \pm 0.09) = 1.19^{+0.14}_{-0.12}$ . It disagrees with our result and the aforementioned papers. We found that our BCG identifications are the same as them in the common 35 clusters and the average relationship of the BCG masses with the Salpeter IMF is  $M_{\text{BCG}}^{\text{Lavoie}} = (1.09 \pm 0.07) M_{\text{BCG}}$ . Therefore, this discrepancy is likely caused by the different fitting methods. Since they used the BCG mass as the  $x$  quantity and ignored their measurement error at each fitting run, their underestimation of the slope ( $1/\beta_{\text{BCG}} = 0.84$ ) in their fitting method could be caused by the regression dilution effect (Appendix A and B). We also mention that the  $\chi^2$  minimization using the likelihood function without the determinant term is not adequate for scaling relation analysis because the determinant (eq. A2) includes the parameters of the slopes and intrinsic scatter and one cannot ignore it (see details; Okabe et al. 2010). We confirmed to recover their steep slope by their  $\chi^2$  minimization method that they randomly pick up values from the normal distribution with a mean of the observables and a standard deviation of the measurement errors. When we properly treat the measurement errors and the regression dilution effect in our code, we find  $\beta_{\text{BCG}} = 0.72^{+0.24}_{-0.17}$  for their sample. The value is still steeper but does not conflict with our result.

Ricci et al. (2018) studied the luminosity function of 142 XXL selected clusters and found by the  $\chi^2$  minimization that the median BCG magnitude is brighter with both redshift and richness. The median BCG luminosity is proportional to  $(1+z)^{1.12 \pm 0.28}$  and  $\lambda^{0.24 \pm 0.08}$ , recomputed from Table 4 of Ricci et al. (2018), where  $\lambda$  is the cluster richness. Although it is difficult to fairly compare with our results because the observables are not the same, the shallow slope of mass dependence is similar to our results and a difference of redshift evolution is only  $\sim 2\sigma$ .

#### 4.4.3 Intrinsic covariance

Farahi et al. (2019) found an anti-correlation between intrinsic scatter of the gas mass and the K-band luminosity of galaxies from 41 LoCuSS clusters of average mass  $M_{500} \sim 4 \times 10^{14} M_{\odot}$ . The pairwise correlation coefficient is  $-0.56^{+0.36}_{-0.28}$ . We did not find significant negative correlation coefficient at  $M_{500} \gtrsim 10^{13} M_{\odot}$ . We compute the probability of accidental anti-correlation for their 41 clusters and find  $\mathcal{P}(r \geq |r_{*,g}|) < 0.2$ , which is not significant. We compute a probability to accidentally realize their correlation coefficient using our 136 clusters, and obtain  $< 2 \times 10^{-2}$ . However, we do not find such a negative correlation. Since we cannot rule out a possibility that the intrinsic correlation depends on halo mass, we need a larger sample of clusters for a more precise measurement.

To our knowledge, the positive correlation coefficients between the BCG and total stellar mass and between the stellar mass of BCG and satellite galaxies (Sec. 3.5) have not been

reported in previous observational studies.

## 5 Summary

We carried out Bayesian analysis for the multivariate scaling relations of the baryonic components of the 136 XXL groups and clusters over a wide range of nearly two orders of magnitude in mass at  $0 \lesssim z \lesssim 1$ . We combined the HSC-SSP weak-lensing mass measurements (Umetsu et al. 2020), the XXL X-ray gas measurements, and the HSC-SSP and SDSS photometry. Bayesian regression simultaneously and consistently takes into account weak-lensing mass calibration, selection effect, error covariance matrix, and intrinsic covariance. The analysis constrains well the slopes, normalizations, and intrinsic covariance among the baryonic component masses. Our method models the parent population together with the scaling relations. Thanks to this modeling, we can correct for selection effect and regression dilution effect. The method does not deal with cosmological inference and it is not affected by uncertainties in the estimation of, e.g.  $\sigma_8$ .

The slope of the gas mass and cluster mass scaling relation is  $1.29^{+0.16}_{-0.10}$  steeper than predicted by the self-similar model ( $\beta = 1$ ), while the slope of the stellar mass is  $0.85^{+0.12}_{-0.09}$  shallower (Table 1, Figures 5 and 11, and Sec. 3.1). As shown in Figure 5 and Sec. 3.3, the gas mass fraction increases from  $f_g(\Omega_b/\Omega_m)^{-1} \sim 0.3$  at  $M_{500}E(z) \sim 10^{13} M_{\odot}$  to  $\sim 0.9$  at  $\sim 10^{15} M_{\odot}$ . The stellar mass fraction decreases from  $f_*(\Omega_b/\Omega_m)^{-1} \sim 0.2$  at  $M_{500}E(z) \sim 10^{13} M_{\odot}$  to  $\sim 0.1$  at  $\sim 10^{15} M_{\odot}$ . Accordingly, the total baryon fraction increases  $f_b(\Omega_b/\Omega_m)^{-1} \sim 0.5$  at  $M_{500}E(z) \sim 10^{13} M_{\odot}$  to  $\sim 0.6$  at  $\sim 10^{14} M_{\odot}$  and  $\sim 1.0$  at  $\sim 10^{15} M_{\odot}$ . The low baryon fraction implies that clusters in our mass range are likely to be open-box systems. The baryon, gas mass, and stellar mass fractions as a function of  $M_{500}$  agree with previous numerical simulations and some previous observations (Figure 12 and Sec 4.3.1 and 4.4.1). Our analysis can differ from previous works for the treatment of the aperture radius and the mass calibration. The slope of the BCG stellar mass is  $0.49^{+0.11}_{-0.10}$  shallower than the other two slopes, indicating a weak-mass dependence (Table 1 and Sec. 3.1). We do not find a significant evidence of redshift evolution in the scaling relations (Table 5 and Sec 3.8) because of their large errors.

The intrinsic scatter is ranked as  $\sigma_g < \sigma_* < \sigma_{\text{BCG}}$ , as numerical simulations (Pillepich et al. 2018; Farahi et al. 2018a). We found a positive intrinsic correlation coefficient between stellar mass and BCG stellar mass (Table 2 and Sec 3.5), which is statistically significant, in agreement with numerical simulations (Anbajagane et al. 2020; Farahi et al. 2020). The intrinsic correlation between gas and stellar mass shows no positive nor negative correlation (Table 2 and Sec 3.5), but the statistical significance is marginal.

We do not find a significant difference between the clusters with and without central radio sources (Table 3, Figure 8, and Sec 3.6). It implies that the cumulative quantity such as the gas and stellar masses are insensitive to the current radio AGN activity. The intrinsic scatter of the total and BCG stellar masses in the C2 clusters with lower mass is larger than that of the C1 clusters with higher mass (Table 4).

This paper comprises the largest sample of X-ray selected clusters over a wide range of nearly two orders of magnitude in mass. Studies of the mass-dependence of the intrinsic covariance and the local slope are out of scope and they could be at reach of the final sample of the XXL clusters with the XXL selection function depending on the cosmological parameters. Studies on subsamples divided by mergers (e.g. Okabe et al. 2019, 2021) will be also done. Analyses of optically-selected (e.g. Rykoff et al. 2014; Oguri 2014; Rozo et al. 2016; Oguri et al. 2018; Maturi et al. 2019), SZ (e.g. Planck Collaboration et al. 2014; Hilton et al. 2021; Bleem et al. 2021) and shear-selected (e.g. Miyazaki et al. 2018a; Oguri et al. 2021) clusters with more than 100 sample sizes are essential to understand the selection effect of cluster finders on baryonic physics.

## Acknowledgments

We gratefully thank the anonymous referee for careful reading and helpful comments. The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at <http://dm.lsst.org>

This paper is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ. We are honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

XXL is an international project based around an XMM Very Large Programme surveying two 25 deg<sup>2</sup> extragalactic fields at a depth of  $\sim 6 \times 10^{-15}$  erg cm<sup>-2</sup> s<sup>-1</sup> in the [0.5-2] keV band for point-like sources. The XXL website is <http://irfu.cea.fr/xxl>. Multi-band information and spectroscopic follow-up of the X-ray sources are obtained through a number of survey programmes, summarised at <http://xxlmultiwave.pbworks.com/>.

This work was supported by Core Research for Energetic Universe in Hiroshima University (the MEXT program for promoting the enhancement of research universities, Japan). MS acknowledges financial contribution from contract ASI-INAF n.2017-14-H.0 and INAF ‘Call per interventi aggiuntivi a sostegno della ricerca di main stream di INAF’. K.U. acknowledges support from the Ministry of Science and Technology of Taiwan (grants MOST 106-2628-M-001-003-MY3 and MOST 109-2112-M-001-018-MY3) and from the Academia Sinica Investigator Award (grant AS-IA-107-M01). SE acknowledges financial contribution from the contracts ASI-INAF Athena 2019-27-HH.0, ‘Attività di Studio per la comunità scientifica di Astrofisica delle Alte Energie e Fisica Astroparticellare’ (Accordo Attuativo ASI-INAF n. 2017-14-H.0), INAF main-stream project 1.05.01.86.10, and from the European Union’s Horizon 2020 Programme under the AHEAD2020 project (grant agreement n. 871158).

## References

- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2004, *AJ*, 128, 502
- Abell, G. O., Corwin, Harold G., J., & Olowin, R. P. 1989, *ApJS*, 70, 1
- Adami, C., Giles, P., Koulouridis, E., et al. 2018, *A&A*, 620, A5, (XXL Paper XX)
- Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, *ApJS*, 240, 23
- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, *ApJS*, 249, 3

- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018a, PASJ, 70, S8
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018b, PASJ, 70, S4
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
- Akras, M. G., & Bershad, M. A. 1996, ApJ, 470, 706
- Allgood, B., Flores, R. A., Primack, J. R., et al. 2006, MNRAS, 367, 1781
- Anbajagane, D., Evrard, A. E., Farahi, A., et al. 2020, MNRAS, 495, 686
- Anders, E., & Grevesse, N. 1989, Geochim. Cosmochim. Acta, 53, 197
- Barnes, D. J., Kay, S. T., Henson, M. A., et al. 2017a, MNRAS, 465, 213
- Barnes, D. J., Kay, S. T., Bahé, Y. M., et al. 2017b, MNRAS, 471, 1088
- Becker, M. R., & Kravtsov, A. V. 2011, ApJ, 740, 25
- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
- Bellstedt, S., Lidman, C., Muzzin, A., et al. 2016, MNRAS, 460, 2862
- Bleem, L. E., Crawford, T. M., Ansarinejad, B., et al. 2021, arXiv e-prints, arXiv:2102.05033
- Böhringer, H., Schuecker, P., Guzzo, L., et al. 2004, A&A, 425, 367
- Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5
- Bradshaw, C., Leauthaud, A., Hearin, A., Huang, S., & Behroozi, P. 2020, MNRAS, 493, 337
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Chabrier, G. 2003, PASP, 115, 763
- Chiu, I., Mohr, J., McDonald, M., et al. 2016a, MNRAS, 455, 258
- Chiu, I., Saro, A., Mohr, J., et al. 2016b, MNRAS, 458, 379
- Chiu, I., Mohr, J. J., McDonald, M., et al. 2018, MNRAS, 478, 3072
- Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8
- Cooper, A. P., Gao, L., Guo, Q., et al. 2015, MNRAS, 451, 2703
- Coupon, J., Czakon, N., Bosch, J., et al. 2018, PASJ, 70, S7
- De Propriis, R., Colless, M., Peacock, J. A., et al. 2004, MNRAS, 351, 125
- Decker, B., Brodwin, M., Abdulla, Z., et al. 2019, ApJ, 878, 72
- DeMaio, T., Gonzalez, A. H., Zabludoff, A., et al. 2020, MNRAS, 491, 3751
- Ebeling, H., Edge, A. C., Allen, S. W., et al. 2000, MNRAS, 318, 333
- Ebeling, H., Edge, A. C., Bohringer, H., et al. 1998, MNRAS, 301, 881
- Eckert, D., Finoguenov, A., Ghirardini, V., et al. 2020, The Open Journal of Astrophysics, 3, 12
- Eckert, D., Ettori, S., Coupon, J., et al. 2016, A&A, 592, A12, (XXL Paper XIII)
- Erben, T., Hildebrandt, H., Miller, L., et al. 2013, MNRAS, 433, 2545
- Erfanianfar, G., Finoguenov, A., Furnell, K., et al. 2019, A&A, 631, A175
- Ettori, S. 2003, MNRAS, 344, L13
- . 2015, MNRAS, 446, 2629
- Evrard, A. E. 1997, MNRAS, 292, 289
- Faccioli, L., Pacaud, F., Sauvageot, J.-L., et al. 2018, A&A, 620, A9, (XXL Paper XXIV)
- Farahi, A., Evrard, A. E., McCarthy, I., Barnes, D. J., & Kay, S. T. 2018a, MNRAS, 478, 2618
- Farahi, A., Ho, M., & Trac, H. 2020, MNRAS, 493, 1361
- Farahi, A., Guglielmo, V., Evrard, A. E., et al. 2018b, A&A, 620, A8, (XXL Paper XXIII)
- Farahi, A., Mulroy, S. L., Evrard, A. E., et al. 2019, Nature Communications, 10, 2504
- Furusawa, H., Koike, M., Takata, T., et al. 2018, PASJ, 70, S3
- Giles, P. A., Maughan, B. J., Pacaud, F., et al. 2016, A&A, 592, A3, (XXL Paper III)
- Gonzalez, A. H., Sivanandam, S., Zabludoff, A. I., & Zaritsky, D. 2013, ApJ, 778, 14
- Guglielmo, V., Poggianti, B. M., Vulcani, B., et al. 2018, A&A, 620, A7, (XXL Paper XXII)
- Hamana, T., Oguri, M., Shirasaki, M., & Sato, M. 2012, MNRAS, 425, 2287
- Henden, N. A., Puchwein, E., & Sijacki, D. 2020, MNRAS, 498, 2114
- Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
- Hilton, M., Sifón, C., Naess, S., et al. 2021, ApJS, 253, 3
- Hoekstra, H. 2003, MNRAS, 339, 1155
- Huang, S., Leauthaud, A., Greene, J. E., et al. 2018a, MNRAS, 475, 3348
- Huang, S., Leauthaud, A., Murata, R., et al. 2018b, PASJ, 70, S6
- Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78
- Jing, Y. P., & Suto, Y. 2002, ApJ, 574, 538
- Kawanomoto, S., Uruguchi, F., Komiyama, Y., et al. 2018, PASJ, 70, 66
- Kelly, B. C. 2007, ApJ, 665, 1489
- Komiyama, Y., Chiba, M., Tanaka, M., et al. 2018, ApJ, 853, 29
- Kravtsov, A. V., Nagai, D., & Vikhlinin, A. A. 2005, ApJ, 625, 588
- Kravtsov, A. V., Vikhlinin, A. A., & Meshcheryakov, A. V. 2018, Astronomy Letters, 44, 8
- Laganá, T. F., Martinet, N., Durret, F., et al. 2013, A&A, 555, A66
- Lavoie, S., Willis, J. P., Démoclès, J., et al. 2016, MNRAS, 462, 4141, (XXL Paper XV)
- Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, MNRAS, 441, 1270
- . 2017, MNRAS, 466, 4442
- Lidman, C., Suherli, J., Muzzin, A., et al. 2012, MNRAS, 427, 550
- Lieu, M., Smith, G. P., Giles, P. A., et al. 2016, A&A, 592, A4, (XXL Paper IV)
- Lin, Y.-T., Mohr, J. J., & Stanford, S. A. 2003, ApJ, 591, 749
- . 2004, ApJ, 610, 745
- Lin, Y.-T., Stanford, S. A., Eisenhardt, P. R. M., et al. 2012, ApJ, 745, L3
- Lin, Y.-T., Hsieh, B.-C., Lin, S.-C., et al. 2017, ApJ, 851, 139
- Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. J. Harnden, F. R., F. A. Primini, & H. E. Payne, 269
- Mandelbaum, R., Miyatake, H., Hamana, T., et al. 2018a, PASJ, 70, S25
- Mandelbaum, R., Lanusse, F., Leauthaud, A., et al. 2018b, MNRAS, 481, 3170
- Mantz, A., Allen, S. W., Ebeling, H., Rapetti, D., & Drlica-Wagner, A. 2010, MNRAS, 406, 1773
- Mantz, A. B. 2019, MNRAS, 485, 4863
- Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2016, MNRAS, 463, 3582
- Martizzi, D., Mohammed, I., Teyssier, R., & Moore, B. 2014, MNRAS, 440, 2290
- Maturi, M., Bellagamba, F., Radovich, M., et al. 2019, MNRAS, 485, 498
- McCarthy, I. G., Bird, S., Schaye, J., et al. 2018, MNRAS, 476, 2999
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936
- McCarthy, I. G., Schaye, J., Bower, R. G., et al. 2011, MNRAS, 412, 1965
- Medezinski, E., Oguri, M., Nishizawa, A. J., et al. 2018, PASJ, 70, 30
- Miyazaki, S., Hamana, T., Ellis, R. S., et al. 2007, ApJ, 669, 714
- Miyazaki, S., Oguri, M., Hamana, T., et al. 2018a, PASJ, 70, S27
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018b, PASJ, 70, S1
- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
- Mulroy, S. L., Farahi, A., Evrard, A. E., et al. 2019, MNRAS, 484, 60
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18

- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Nishizawa, A. J., Hsieh, B.-C., Tanaka, M., & Takata, T. 2020, arXiv e-prints, arXiv:2003.01511
- Nishizawa, A. J., Oguri, M., Oogi, T., et al. 2018, *PASJ*, 70, S24
- Oguri, M. 2014, *MNRAS*, 444, 147
- Oguri, M., & Hamana, T. 2011, *MNRAS*, 414, 1851
- Oguri, M., Lin, Y.-T., Lin, S.-C., et al. 2018, *PASJ*, 70, S20
- Oguri, M., Miyazaki, S., Li, X., et al. 2021, *PASJ*, 73, 817
- Okabe, N., & Smith, G. P. 2016, *MNRAS*, 461, 3794
- Okabe, N., Zhang, Y.-Y., Finoguenov, A., et al. 2010, *ApJ*, 721, 875
- Okabe, N., Umetsu, K., Tamura, T., et al. 2016, *MNRAS*, 456, 4475
- Okabe, N., Oguri, M., Akamatsu, H., et al. 2019, *PASJ*, 71, 79
- Okabe, N., Dicker, S., Eckert, D., et al. 2021, *MNRAS*, 501, 1701
- Pacaud, F., Pierre, M., Refregier, A., et al. 2006, *MNRAS*, 372, 578
- Pacaud, F., Clerc, N., Giles, P. A., et al. 2016, *A&A*, 592, A2, (XXL Paper II)
- Pierre, M., Pacaud, F., Adami, C., et al. 2016, *A&A*, 592, A1, (XXL Paper I)
- Piffaretti, R., Arnaud, M., Pratt, G. W., Pointecouteau, E., & Melin, J.-B. 2011, *A&A*, 534, A109
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, 475, 648
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A29
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Planelles, S., Borgani, S., Dolag, K., et al. 2013, *MNRAS*, 431, 1487
- Pompei, E., Adami, C., Eckert, D., et al. 2016, *A&A*, 592, A6, (XXL Paper VII)
- Pozzetti, L., Bolzonella, M., Lamareille, F., et al. 2007, *A&A*, 474, 443
- Pratt, G. W., Arnaud, M., Biviano, A., et al. 2019, *Space Sci. Rev.*, 215, 25
- Ricci, M., Benoist, C., Maurogordato, S., et al. 2018, *A&A*, 620, A13, (XXL Paper XXVIII)
- Rousseeuw, P. J., & Molenberghs, G. 1994, *The American Statistician*, 48, 276
- Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, *MNRAS*, 461, 1431
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, *ApJ*, 785, 104
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, *PyMC3: Python probabilistic programming framework*
- Schechter, P. 1976, *ApJ*, 203, 297
- Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, *A&A*, 609, A84
- Sembolini, F., Yepes, G., Pearce, F. R., et al. 2016a, *MNRAS*, 457, 4063
- Sembolini, F., Elahi, P. J., Pearce, F. R., et al. 2016b, *MNRAS*, 459, 2973
- Sereno, M. 2016, *MNRAS*, 455, 2149
- Sereno, M., Umetsu, K., Ettori, S., et al. 2020, *MNRAS*, 492, 4528
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, 214, 24
- Smith, R. K., Brickhouse, N. S., Liedahl, D. A., & Raymond, J. C. 2001, *ApJ*, 556, L91
- Tanaka, M. 2015, *ApJ*, 801, 20
- Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, *PASJ*, 70, S9
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, 688, 709
- Truong, N., Rasia, E., Mazzotta, P., et al. 2018, *MNRAS*, 474, 4089
- Umetsu, K. 2020, *A&A Rev.*, 28, 7
- Umetsu, K., Sereno, M., Lieu, M., et al. 2020, *ApJ*, 890, 148
- Valotti, A., Pierre, M., Farahi, A., et al. 2018, *A&A*, 614, A72
- Vikhlinin, A., Burenin, R. A., Ebeling, H., et al. 2009, *ApJ*, 692, 1033
- White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, *ApJ*, 475, 479
- White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, *Nature*, 366, 429
- Whitmore, B. C., Gilmore, D. M., & Jones, C. 1993, *ApJ*, 407, 489
- Willis, J. P., Oguri, M., Ramos-Ceja, M. E., et al. 2021, *MNRAS*, 503, 5624
- Wu, H.-Y., Evrard, A. E., Hahn, O., et al. 2015, *MNRAS*, 452, 1982
- Young, O. E., Thomas, P. A., Short, C. J., & Pearce, F. 2011, *MNRAS*, 413, 691
- Zhang, Y., Miller, C., McKay, T., et al. 2016, *ApJ*, 816, 98
- Zhang, Y.-Y., Okabe, N., Finoguenov, A., et al. 2010, *ApJ*, 711, 1033

## Appendix A Linear regression for multivariate scaling relations

Let us consider an observed dataset  $\{x_n, \mathbf{y}_n\}_{n=1}^N$ , composed of  $D + 1$  variables for  $N$  sampled clusters, where  $\mathbf{y} = \{y_0, y_1, \dots, y_D\}$ . The true variables of  $\{X_n, \mathbf{Y}_n\}_{n=1}^N$  ( $\mathbf{Y} = \{Y_0, Y_1, \dots, Y_D\}$ ) for the  $n$ -th cluster sample are related to the observables by  $p(x_n, \mathbf{y}_n | X, \mathbf{Y}) = \mathcal{N}(\{x_n, \mathbf{y}_n\} | \{X_n, \mathbf{Y}_n\}, \mathbf{\Sigma}_{\text{err},n})$  with Gaussian error covariance matrix,  $\mathbf{\Sigma}_{\text{err}}$ . In cluster scaling relation studies, the  $x$  value can be a logarithm on the weak-lensing, hydrostatic mass or mass estimates from a scaling relation. They are affected by scatter and/or bias from the true mass because of some observational systematics (e.g. Becker & Kravtsov 2011; Zhang et al. 2010; Hamana et al. 2012; Okabe & Smith 2016; Pratt et al. 2019; Umetsu et al. 2020). Following Sereno (2016), we introduce the latent variable  $Z$  as the true mass (see also Mulroy et al. 2019; Farahi et al. 2019). The linear regression equation between  $\{X, \mathbf{Y}\}$  and  $Z$  are expressed by  $X_Z = \alpha_X + \beta_X Z$  and  $\mathbf{Y}_Z = \boldsymbol{\alpha} + \beta \mathbf{Z}$ , respectively. Given the intrinsic covariance matrix ( $\mathbf{\Sigma}_{\text{int}}$ , e.g. Okabe et al. 2010), the conditional probability,  $p(X_n, \mathbf{Y}_n | Z_n, \boldsymbol{\theta})$ , is a multivariate normal density distribution of  $p(X_n, \mathbf{Y}_n | Z_n, \boldsymbol{\theta}) = \mathcal{N}(\{X_n, \mathbf{Y}_n\} | \{\alpha_X + \beta_X Z_n, \boldsymbol{\alpha} + \beta \mathbf{Z}_n\}, \mathbf{\Sigma}_{\text{int}})$ , where  $\boldsymbol{\theta}$  denotes the parameters of the distribution. The element of the intrinsic covariance matrix is specified by

$$\mathbf{\Sigma}_{\text{int}} = \begin{pmatrix} \sigma_X^2 & r_{XY_i} \sigma_X \sigma_{Y_i} & r_{XY_j} \sigma_X \sigma_{Y_j} \\ r_{XY_i} \sigma_X \sigma_{Y_i} & \sigma_{Y_i}^2 & r_{Y_i Y_j} \sigma_{Y_i} \sigma_{Y_j} \\ r_{XY_j} \sigma_X \sigma_{Y_j} & r_{Y_i Y_j} \sigma_{Y_i} \sigma_{Y_j} & \sigma_{Y_j}^2 \end{pmatrix} \quad (\text{A1})$$

where  $\sigma_X$  and  $\sigma_{Y_i}$  are the intrinsic scatter and  $r_{XY_i}$  and  $r_{Y_i Y_j}$  are the intrinsic correlation coefficient between  $X$  and  $Y_i$  and between  $Y_i$  and  $Y_j$ , respectively. All the elements in the intrinsic covariance should satisfy the condition of covariance correlation matrix: all the eigenvalues are positive. Since the intrinsic scatter is a positive quantity, we use logarithmic quantities,  $\ln \sigma_X$  and  $\ln \sigma_{Y_i}$ , in actual computations to avoid boundary artifacts at zero.

Linear regression of multi-wavelength datasets from a survey must take into account for two systematic effects. First, the slopes can be underestimated by the measurement errors of the  $x$  value, so-called, regression dilution effect (Akritas & Bershady 1996; Kelly 2007; Sereno 2016). Second, selection effects, e.g. Malmquist bias, affects sample selected above an observational threshold (e.g. Mantz et al. 2010; Sereno 2016; Mantz et al. 2016; Mulroy et al. 2019; Mantz 2019). We follow the mathematical formulation of Sereno (2016) to overcome the above two problems. We introduce the parent population,  $p(Z | \boldsymbol{\theta})$ , assuming the Gaussian distribution,  $\mathcal{N}(\mu_Z, \sigma_Z)$ , where  $\mu_Z$  and  $\sigma_Z$  are hyper-parameters. In a generalized case, the parent population can be the summation of multiple Gaussian distributions:  $p(Z | \boldsymbol{\theta}) = \sum_i \pi_i \mathcal{N}(\mu_{Z,i}, \sigma_{Z,i})$ , where  $\pi_i$  is the

normalization satisfying with  $\sum_i \pi_i = 1$ . The standard deviation,  $\sigma_Z$ , effectively corrects for the regression dilution effect (Akritas & Bershady 1996; Kelly 2007; Sereno 2016). The total parameters are  $\boldsymbol{\theta} = \alpha, \beta, \mathbf{\Sigma}_{\text{int}}, \mu_Z, \sigma_Z$ . The selection bias is modelled by truncating the probability distribution with the threshold of  $y_{\text{th},0,n}$  on a tracer  $y_0$  for cluster finders, where the subscript  $n$  denotes the  $n$ -th cluster. The Bayesian chain rule gives the likelihood function of the  $D$ -dimensional scaling relations, as follows,

$$p(x, \mathbf{y} | \boldsymbol{\theta}) = \frac{\prod_n \int_{-\infty}^{\infty} dX_n \int_{-\infty}^{\infty} d^D \mathbf{Y}_n \int_{-\infty}^{\infty} dZ_n p(x_n, \mathbf{y}_n | X_n, \mathbf{Y}_n) p(X_n, \mathbf{Y}_n | Z_n, \boldsymbol{\theta}) p(Z_n | \boldsymbol{\theta})}{\prod_n \int_{y_{\text{th},0,n}}^{\infty} dy_{0,n} \int_{-\infty}^{\infty} dx_n \int_{-\infty}^{\infty} d^{D-1} \mathbf{y}_n \int_{-\infty}^{\infty} dX_n \int_{-\infty}^{\infty} d^D \mathbf{Y}_n \int_{-\infty}^{\infty} dZ_n p(x_n, \mathbf{y}_n | X_n, \mathbf{Y}_n) p(X_n, \mathbf{Y}_n | Z_n, \boldsymbol{\theta}) p(Z_n | \boldsymbol{\theta})} \quad (\text{A2})$$

where  $n$  denotes the  $n$ -th cluster,  $d^{D-1} \mathbf{y}_n = dy_{n,1} dy_{n,2} \cdots dy_{n,D}$ , and  $d^D \mathbf{Y}_n = dY_{n,0} dY_{n,1} \cdots dY_{n,D}$ . Given Bayes' theorem, a conditional probability given the observables is expressed by  $p(\boldsymbol{\theta} | x, \mathbf{y}) \propto p(x, \mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$  where  $p(\boldsymbol{\theta})$  is the prior distribution of the parameters. The method takes into account the intrinsic covariance between cluster properties ( $\Sigma_{\text{int}}$ ), selection effect ( $\sigma_Z, \mu_Z, y_{\text{th},0}$ ), weak-lensing mass calibration ( $\alpha_X, \beta_X, \sigma_X$ ), and observational error covariance matrix ( $\Sigma_{\text{err}}$ ), for the purpose of the analysis of the multivariate scaling relations.

We adopt the Markov Chain Monte Carlo (MCMC) method and a biweight estimate of the posterior distributions. We use a flat prior  $[-10^4, 10^4]$  on  $\boldsymbol{\theta} = \boldsymbol{\alpha}$ , each parameter of the intrinsic covariance of  $\Sigma_{\text{int}}$  (eq. A1), and  $\mu_Z$ . We employ a Student's  $t_1$  distribution with one degree of freedom on  $\beta$  so that slope angles become uniformly distributed. A non-informative prior distribution on the variance of the parent distribution,  $\sigma_Z^2$ , follows a scaled inverse  $\chi^2$  distribution as a conjugate prior satisfying that posterior distributions have the same probability distribution family as the prior distribution.

When we consider redshift evolution in linear regression and  $p(Z)$ , the linear regression forms  $\alpha_X + \beta_X Z + \gamma_X \ln E(z)$  and  $\alpha + \beta Z + \gamma \ln E(z)$  and the parameters of  $p(Z, z)$  are described by  $\mu_Z(z) = \mu_{Z,0} + \gamma_{\mu_Z} \ln E(z)$  and  $\ln \sigma_Z(z) = \ln \sigma_{Z,0} + \gamma_{\sigma_Z} \ln E(z)$ . Since we assume no error in cluster redshifts, the regression dilution effect disappears. The prior of redshift-dependent slopes,  $\gamma$ , is a Student's  $t_1$  distribution with one degree of freedom.

There is also a practical problem of numerical errors in computing the inverse of the composition matrix of the intrinsic covariance and the measurement errors. As aforementioned, all the parameters in the intrinsic covariance and the composition matrix should satisfy the condition of the covariance correlation matrix. We compute the eigenvalues of the two matrices by the parameters randomly and independently proposed in each step. If there is at least one negative eigenvalue, we re-propose new parameters. Otherwise, the inverse matrix calculation has large numerical errors especially in higher dimensions, giving poorly-constrained posterior distributions.

Figure 13 demonstrates the regression dilution effect and the selection effect in 1D scaling relations from mock simulations including errors and intrinsic scatter. The left panel uses data independent of any selection processes and the right panel uses a sample catalog selected from a mass function through a mass-

observable scaling relation with the truncated threshold on the  $y$  quantity. The code recovers well the input parameters, while the computation without  $p(Z | \boldsymbol{\theta})$  (regression dilution effect) or with  $y_{\text{th},0,n} \rightarrow -\infty$  (selection effect) underestimates the slopes.

We further assess the performance of the newly developed code using eight mock simulations, each of which is composed of 500 realizations. The first six simulations use a 2D or 4D scaling relation with different setup parameters. Individual measurement errors in  $\{x, \mathbf{y}\}$  are randomly assigned from  $\sigma^{\text{err}} = [0.05, 0.2]$  or  $\sigma^{\text{err}} = [0.4, 0.6]$ , we refer them to as 10% and 50%, respectively. We assumed two cases with the error correlation  $r^{\text{err}} = 0$  or  $r^{\text{err}} = [-1, 1]$ . Each cluster is drawn through a tracer  $y_0$  from a parent population of Gaussian distribution or a halo mass function of Tinker et al. (2008) at  $z = 0.3$ . Each parameter combination are referred in the  $x$  label of Fig. 14. We simulate  $\sim 100$  clusters per run, corresponding to the case of this study. In each simulation, the input parameters of the scaling relations are randomly drawn from a uniform distribution in  $\alpha_Y = [0, 1]$ ,  $\beta_Y = [0.3, 1.7]$ ,  $\sigma_Y = [0.05, 0.4]$ , and  $r_Y = [-0.8, 0.8]$ , respectively. The range of intrinsic scatter,  $\sigma_Y$ , is lower than 50% measurement error, and higher than or comparable to 10% error. We fix external mass calibration parameters;  $\alpha_X = 0$ ,  $\beta_X = 1$ ,  $\sigma_X = 0.1$ , and  $r_{X,i} = 0$ .

We quantify the performance of our code by  $\boldsymbol{\theta}_{\text{output}} = m \times \boldsymbol{\theta}_{\text{input}} + c$ , where  $\boldsymbol{\theta}_{\text{input}}$  is the input parameters,  $\boldsymbol{\theta}_{\text{output}}$  is the output parameter,  $m$  is a multiplicative error, and  $c$  is an additive error. The results are shown in Figure 14. The code can constrain both slope and normalization, irrespective of the parent distribution. While intrinsic scatter and correlation coefficients are constrained well when the measurement errors are smaller than intrinsic scatter, they cannot be constrained when the measurement errors are larger than intrinsic scatter.

We choose the parameters for the last two simulations as similar to measurement errors of this study;  $\sigma_x^{\text{err}} = [0.75, 1.25]$ ,  $\sigma_{y_0}^{\text{err}} = 0.1$ ,  $\sigma_{y_1}^{\text{err}} = [0.2, 0.7]$ ,  $\sigma_{y_2}^{\text{err}} = 0.02$ , and  $\sigma_{y_3}^{\text{err}} = [0.4, 1]$ . The error correlation coefficient is set to be  $r_{x,y_1}^{\text{err}} = 0.85$ ,  $r_{x,y_3}^{\text{err}} = 0.95$ , and  $r_{y_1,y_3}^{\text{err}} = [0.64, 0.97]$  and the others are fixed to zero. The input parameters of the scaling relations are randomly drawn from a uniform distribution in  $\alpha = [0, 1]$ ,  $\beta = [0.3, 1.7]$ ,  $\sigma_{\text{int}} = [0.4, 0.8]$ , and  $r_{\text{int}} = [-0.8, 0.8]$ , respectively. The intrinsic scatter is larger than that employed in the first six simulations. We fix  $\alpha_X = \ln 0.89$ ,  $\beta_X = 1.09$ ,  $\sigma_X = 0.21$ , and  $r_{X,Y_i} = 0$ . As for the parent population, we adopt the two cases of no-redshift dependence  $p(Z)$  and redshift dependence

$p(Z, z)$ . In the latter case, the random redshift is uniformly distributed in  $z = [0.1, 1]$  and the mass function is computed at the given redshift. Then, the sample is constructed with a selection cut similar to the observation (Figure 9). Although the measurement errors for the  $x$  quantity are large, the input parameters, including intrinsic scatter and correlation coefficients, are recovered because the error correlations are all taken into account.

## Appendix B Regression dilution effect

We briefly explain the regression dilution effect (e.g. Akritas & Bershadsky 1996). We first consider the two true variables of  $\{X_n, Y_n\}$  for the  $n$ -th cluster of the sample size  $N$ , where we assume  $\alpha_X = 0$ ,  $\beta_X = 1$ , and  $\sigma_X = 0$  and thus,  $p(X) = p(Z)$ . We assume that the two variables follow a simple linear regression, specified by

$$Y_n = \alpha_Y + \beta_Y X_n + \varepsilon_n, \quad (\text{A3})$$

where  $\varepsilon_n$  is a random variable drawn from the intrinsic scatter  $\mathcal{N}(0, \sigma_Y)$  of the parameter  $Y$ .

The regression parameters,  $\alpha_Y$  and  $\beta_Y$ , are described by the first- and second-order moments of eq. (A3),

$$\alpha_Y = E(Y_n) - \beta_Y E(X_n) = \mu_Y - \beta_Y \mu_X, \quad (\text{A4})$$

$$\beta_Y = \frac{(1/N) \sum_n (X_n - \mu_X)(Y_n - \mu_Y)}{(1/N) \sum_n (X_n - \mu_X)^2} = \frac{\sigma_{XY}}{\sigma_X^2}, \quad (\text{A5})$$

where  $\mu_X = E(X_n)$  and  $\mu_Y = E(Y_n)$  are the average quantities,  $\sigma_X^2$  is a variance of the parent population  $p(X)$ , and  $\sigma_{XY}$  is the covariance of the bivariate  $X$  and  $Y$  distribution.

We next consider a measurement error model on  $X$ ,

$$Y_n = \alpha + \beta x_n + \varepsilon_n \quad (\text{A6})$$

$$x_n = X_n + \eta_n \quad (\text{A7})$$

where  $x_n$  is the observed quantity and  $\eta_n$  is the measurement error drawn from  $\mathcal{N}(0, \sigma_x^{\text{err}})$ , irrespective of  $X_n$ . The regression parameters become

$$\begin{aligned} \beta &= \frac{(1/N) \sum_n (x_n - E(x_n))(Y_n - \mu_Y)}{(1/N) \sum_n (x_n - E(x_n))^2}, \\ &= \frac{(1/N) \sum_n (X_n + \eta_n - \mu_X)(Y_n - \mu_Y)}{(1/N) \sum_n (X_n + \eta_n - \mu_X)^2}, \\ &= \frac{(1/N) \sum_n (X_n - \mu_X)^2 \beta_Y}{(1/N) \sum_n (X_n - \mu_X)^2 + (\sigma_x^{\text{err}})^2}, \\ &= \frac{\sigma_X^2}{\sigma_X^2 + (\sigma_x^{\text{err}})^2} \beta_Y, \end{aligned} \quad (\text{A8})$$

$$\begin{aligned} \alpha &= \mu_Y - \frac{\sigma_X^2}{\sigma_X^2 + (\sigma_x^{\text{err}})^2} \beta_Y \mu_X, \\ &= \alpha_Y + \left(1 - \frac{\sigma_X^2}{\sigma_X^2 + (\sigma_x^{\text{err}})^2}\right) \beta_Y \mu_X \end{aligned} \quad (\text{A9})$$

If the sample size is infinite ( $\sigma_X \rightarrow \infty$ ), the regression parameters,  $\alpha$  and  $\beta$ , coincide with  $\alpha_X$  and  $\beta_X$ , respectively. However,

in the case of the finite sample size, the slope directly estimated by the observables is underestimated by  $(1 + (\sigma_x^{\text{err}})^2 / \sigma_X^2)^{-1}$ . When we consider the measurement error and the variance of the parent population in the regression analysis, we correct the measured value,  $\beta$ , by  $(1 + (\sigma_x^{\text{err}})^2 / \sigma_X^2)$ . The Bayesian forward modeling enables us to directly recover  $\beta_Y$ .

We next introduce the observed quantity  $y_n$ , as follows

$$y_n = \alpha' + \beta' x_n + \varepsilon_n \quad (\text{A10})$$

$$x_n = X_n + \eta_n \quad (\text{A11})$$

$$y_n = Y_n + \xi_n \quad (\text{A12})$$

where  $\xi_n$  is the measurement error of  $y_n$ . The slope parameter becomes

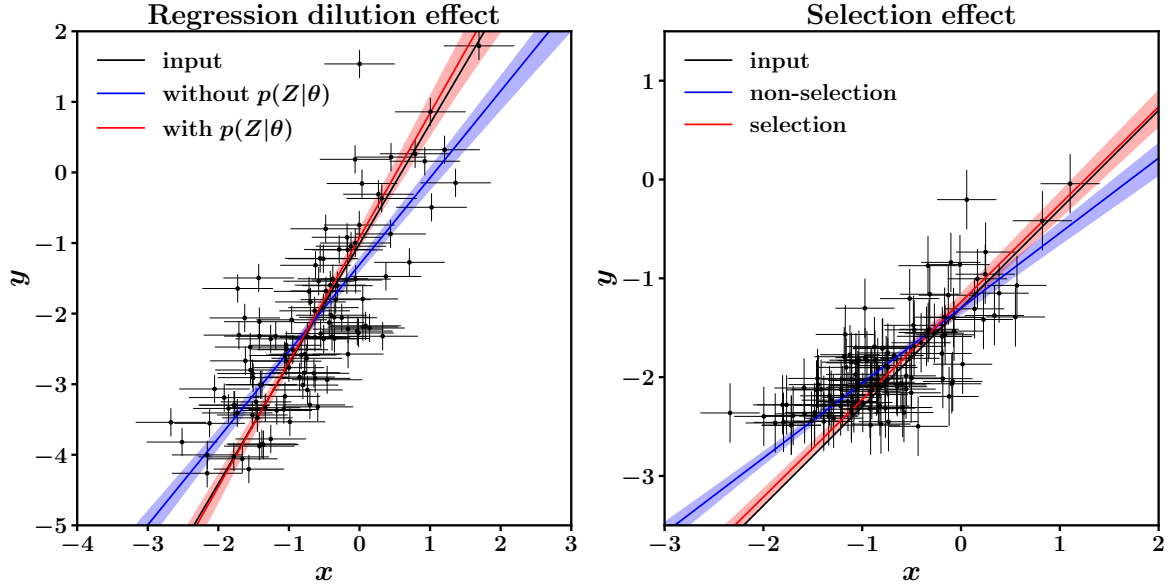
$$\begin{aligned} \beta' &= \frac{(1/N) \sum_n (x_n - E(x_n))(y_n - E(y_n))}{(1/N) \sum_n (x_n - E(x_n))^2}, \\ &= \frac{(1/N) \sum_n (X_n + \eta_n - \mu_X)(Y_n + \xi_n - \mu_Y)}{(1/N) \sum_n (X_n + \eta_n - \mu_X)^2}, \\ &= \frac{(1/N) \sum_n (X_n - \mu_X)(Y_n - \mu_Y) + (1/N) \sum_n \eta_n \xi_n}{(1/N) \sum_n (X_n - \mu_X)^2 + (1/N) \sum_n \eta_n^2}, \\ &= \frac{(1/N) \sum_n (X_n - \mu_X)^2 \beta_Y + \sigma_{xy}^{\text{err}}}{(1/N) \sum_n (X_n - \mu_X)^2 + (\sigma_x^{\text{err}})^2}, \\ &= \frac{\sigma_X^2}{\sigma_X^2 + (\sigma_x^{\text{err}})^2} \beta_Y + \frac{\sigma_{xy}^{\text{err}}}{\sigma_X^2 + (\sigma_x^{\text{err}})^2}, \end{aligned} \quad (\text{A13})$$

where  $\sigma_{xy}^{\text{err}}$  is the error correlation between the  $x$  and  $y$  quantities. If there is no error correlation ( $\sigma_{xy}^{\text{err}} = 0$ ), eq. A13 coincides with eq. A8 and thus the regression dilution effect depends on the error of the  $x$  quantity. If the errors are correlated, it is important to implement the error covariance matrix in the regression analysis (eq. A2).

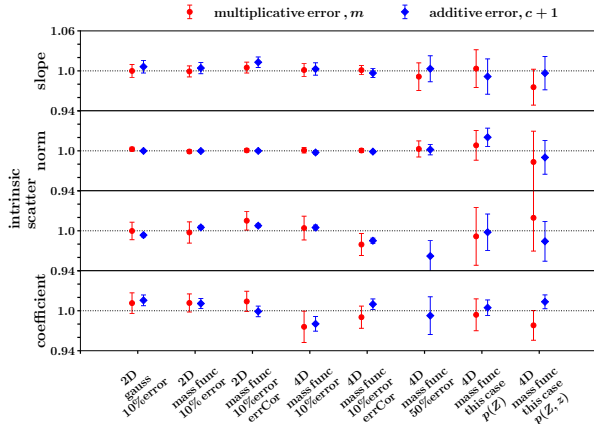
As described above,  $\sigma_x^{\text{err}}$  in the finite sample gives rise to the regression dilution effect, while  $\sigma_y^{\text{err}}$  does not as long as it is not correlated with  $\sigma_x^{\text{err}}$ . Therefore, it is vitally important to infer  $\sigma_X^2$  and  $\mu_X$  to accurately estimate the regression coefficient parameters. Here, we do not assume any specific distributions of the  $X$  quantity, but require only  $\sigma_X^2$  and  $\mu_X$ . In our case, they correspond to  $\sigma_Z^2$  and  $\mu_Z$  of the logarithm distribution of the true mass. Hence, we do not necessarily need a full description of  $p(Z)$  but effective estimations of  $\sigma_X$  and  $\mu$  in the regression analysis. In other words, the introduction of  $\mu_Z$  and  $\sigma_Z$  in our scaling relation analysis is neither for cosmological purposes nor for an accurate measurement of the mass function. This paper assumes the Gaussian distribution,  $\mathcal{N}(\mu_Z, \sigma_Z)$  to correct for the regression dilution effect. The Gaussian distribution enables a fast computation. They are determined as the hyper-parameters to respond flexibly to an unknown population distribution (Appendix A).

## Appendix C Aperture correction

In the above regression analysis, we use the  $Y$  quantities measured within the observed overdensity radii,  $r^X$ , and obtain the



**Fig. 13.** The regression dilution effect (left) and selection effect (right) of the mock simulations. The solid black lines denote the input scaling relations. The solid red lines and transparent regions are the estimate and  $1\sigma$  uncertainty by our code. The solid blue lines and transparent regions are the estimate and  $1\sigma$  uncertainty without a consideration of the regression dilution effect or the selection effect. When we improperly treat these effects, the slopes are underestimated. The data in the right panel is selected above the threshold of  $y = -2.5$ .



**Fig. 14.** A performance of our Bayesian code of slope, normalization, intrinsic scatter and intrinsic correlation coefficient (from top to bottom). A multiplicative error ( $m$ ; red circles) and an additive error ( $c$ ; blue diamonds) in an equation  $\theta_{\text{output}} = m \times \theta_{\text{input}} + c$ . Each error size corresponds to the error per a parameter. We consider 2D and 4D scaling relations with a combination of 10 percent and 50 percent measurement errors in  $\{x, y\}$  and two parent populations. The parent population of samples is assumed to be either a halo mass function (Tinker et al. 2008) at  $z = 0.3$  or Gaussian distribution. Each subset complies 500 simulations. One simulation is composed of  $\sim 100$  clusters. The sample clusters are drawn through a tracer  $y_0$ . Input parameters of scaling relations are uniformly distributed in  $\alpha = [0, 1]$ ,  $\beta = [0.3, 1.7]$ ,  $\sigma_{\text{int}} = [0.05, 0.4]$ , and  $r = [-0.8, 0.8]$ . The code can constrain well slope and normalization, irrespective of a parent distribution. The intrinsic scatter and correlation coefficient can be constrained when the measurement errors are smaller than intrinsic scatter. In the opposite case, neither intrinsic scatter nor correlation coefficient is constrained with a large  $1\sigma$  uncertainty. When we employ the measurement error matrix similar to this study ( $p(Z)$  and  $p(Z, z)$ ), the input parameters are recovered.

scaling relation of  $Y_Z = Y_Z(< r^X) = \alpha + \beta Z$ . When we study the correlation between the true mass and the observables measured within the true mass overdensity,  $r^Z$ , we convert  $Y_Z(< r^X)$  into  $Y_Z(< r^Z)$ . The original baseline,  $Y_Z(< r^X)$ , can be expanded in series up to the first order of a radius deviation,  $\delta r^Z$ , as follows,

$$Y_Z(< r^X) = Y_Z(< r^Z) + \frac{\partial Y_Z(< r^Z)}{\partial r^Z} \delta r^Z. \quad (\text{A14})$$

The new baseline is thus specified by

$$Y_Z(< r^Z) = Y_Z(< r^X) - \frac{\partial Y_Z(< r^Z)}{\partial r^Z} \delta r^Z = \alpha + \delta Y_Z + \beta Z, \quad (\text{A15})$$

where  $\delta Y_Z$  is the aperture correction term in the scaling relation of interest. We consider the logarithmic quantity,  $Y_i \propto \ln \mathcal{O}_i$  and  $Z \propto \ln M(< r^Z)$ , where  $\mathcal{O}_i$  and  $M$  are the observable and the true mass, respectively. Assuming that the averaged density profile for the observable,  $\mathcal{O}_i$ , follows a power law distribution  $\rho_i \propto r^{-p}$ , we find  $\delta Y_Z = -(3-p)(\delta r^Z/r^Z) = -(1-p/3)\delta M(< r^Z)/M(< r^Z)$ . We recall the mass calibration relation between the observed mass and the true mass,  $X = \alpha_X + \beta_X Z$ , and then obtain  $\delta M(< r^Z)/M(< r^Z) = e^{\alpha_X + (\beta_X - 1)Z} - 1$ . In short, the aperture correction term becomes

$$\delta Y_Z = \left(1 - \frac{p}{3}\right) (1 - \exp[\alpha_X + (\beta_X - 1)Z]). \quad (\text{A16})$$

The normalization,  $\alpha + \delta Y_Z$ , in the scaling relation (eq. A15) slightly changes as a function of the true mass.

## Appendix D Posterior distribution

It is worth showing the posterior distributions of our main result (Table 1 and Table 2). Due to the page size limit, we split the 25 parameters into two figures for normalization and slopes (Fig. 15) and intrinsic covariance (Fig. 16) for the  $r_{*,g}^{\text{err}} = r_{\text{WL},*}^{\text{err}} r_{\text{WL},g}^{\text{err}}$  case. Black solid curves represent the trivariate Gaussian distribution prior for the weak-lensing mass calibration (Sec. 2.4.3).

## Appendix E Scaling relations without the $E(z)$ correction

In some numerical simulations (e.g. Martizzi et al. 2014; Le Brun et al. 2014; Wu et al. 2015; McCarthy et al. 2017; Barnes et al. 2017b; Farahi et al. 2018a, 2020) and observational papers (e.g. Lin et al. 2003, 2004; Gonzalez et al. 2013; Laganá et al. 2013), they investigated the scaling relations without the  $E(z)$  correction. It is useful to show the results assuming no redshift evolution. We define the observables independent of redshifts as in eq. 10. The form of scaling relations is similar to eq. 11 except for redshift-dependent slopes,

$$x = \ln \frac{M_{500}^{\text{WL}}}{10^{14} M_{\odot}} \quad (\text{E.1})$$

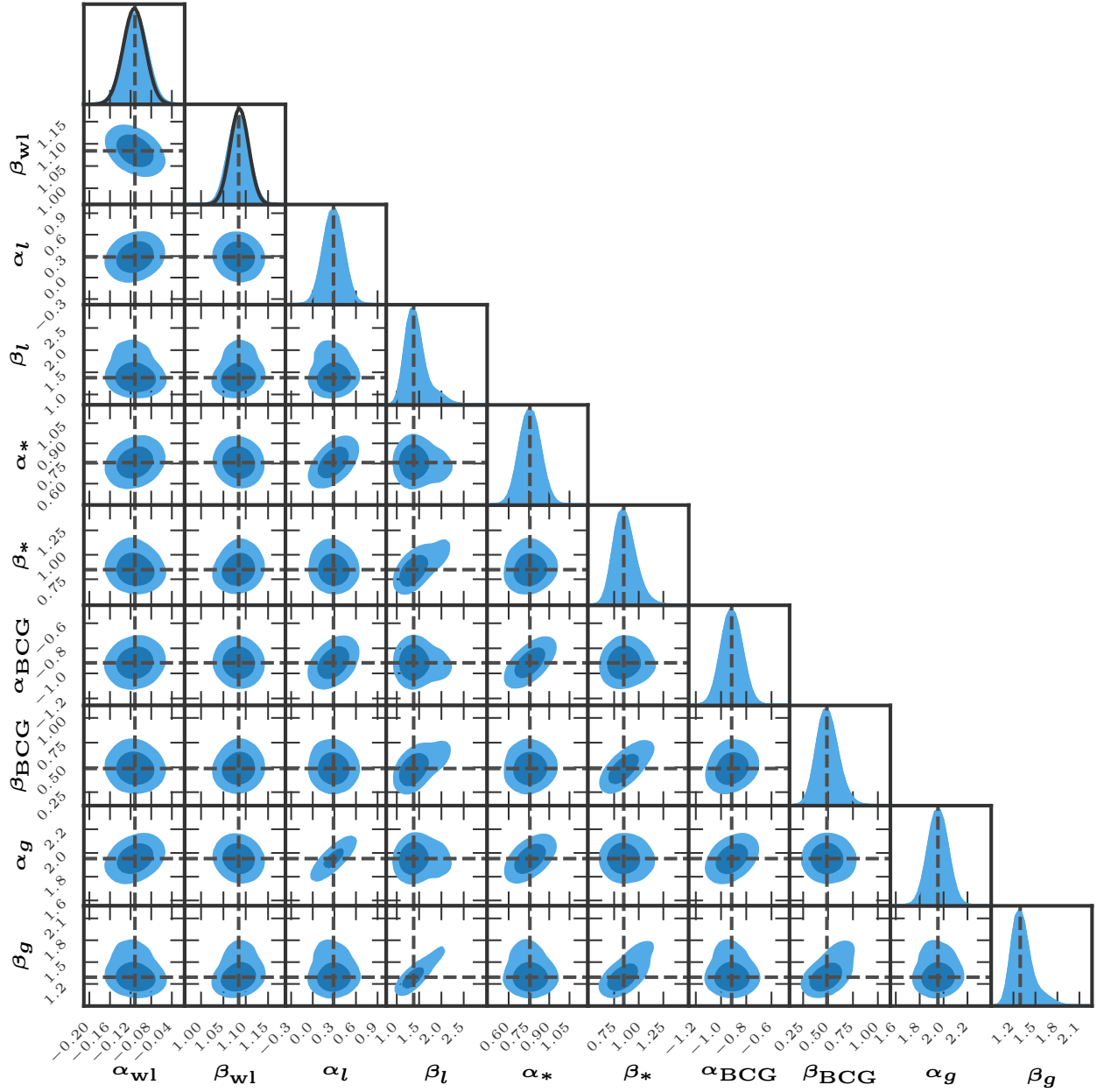
$$\mathbf{y} = \left\{ \ln \frac{L_X}{10^{43} \text{ ergs}^{-1}}, \ln \frac{M_*}{10^{12} M_{\odot}}, \ln \frac{M_{\text{BCG}}}{10^{12} M_{\odot}}, \ln \frac{M_g}{10^{12} M_{\odot}} \right\}$$

$$\mathbf{Y}_Z = \boldsymbol{\alpha} + \beta Z \quad (\text{E.2})$$

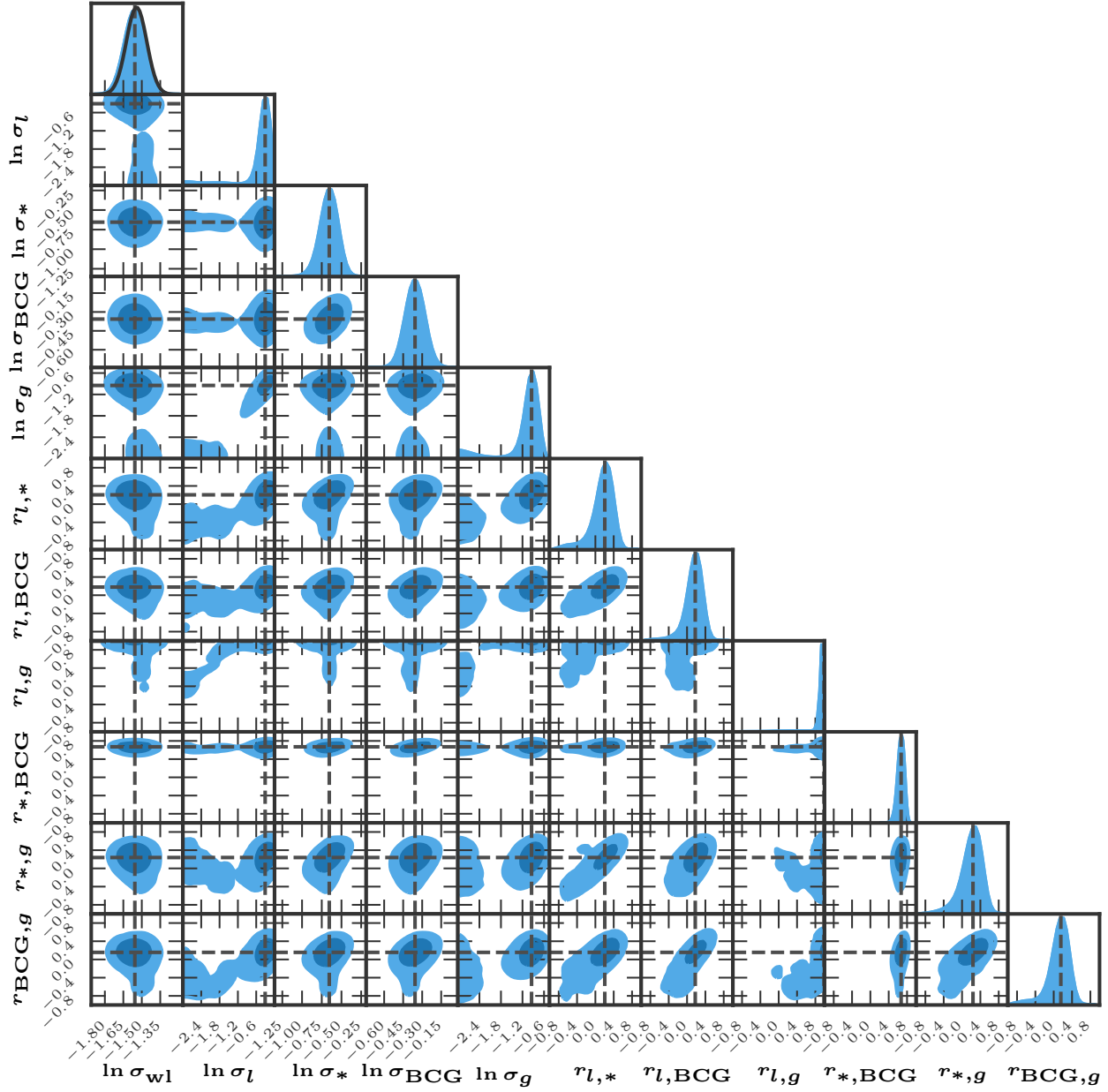
Here, we assume that the scaling relations do not evolve with redshift ( $\gamma = 0$ ), however, we employ the redshift-dependent mean  $\mu_Z(z)$  and standard deviation  $\sigma_Z(z)$  of parent population  $p(Z, z)$  because the sample selection depends on redshift (Sec. 4.1.2 and Figure 9). With this set up, we repeat the Bayesian analysis for scaling relations. The results are shown in Table 7 and Table 8.

**Table 7.** Resulting regression parameters of the scaling relations between the cluster quantities ( $M_{500}^{\text{WL}}$ ,  $L_X$ ,  $M_*$ ,  $M_{\text{BCG}}$ , and  $M_g$ ) and the true mass  $M_{500}$  for the 136 XXL clusters without the  $E(z)$  correction. The normalization,  $\alpha$ , and the slope,  $\beta$ , are defined by the linear regressions (eq. E.2). The intrinsic scatter at a fixed true mass is represented by  $\sigma_{\text{int}}$ . <sup>†</sup>: the results using a trivariate Gaussian prior as the WL mass calibration, as described in Sec. 2.4.3. The errors denote the  $1\sigma$  uncertainty.

	$\alpha$	$\beta$	$\sigma_{\text{int}}$
$M_{500}^{\text{WL}}$	$-0.11^{+0.02}_{-0.02}$	$1.08^{+0.02}_{-0.02}$	$0.21^{+0.02}_{-0.02}$
$L_X$	$0.72^{+0.14}_{-0.43}$	$1.39^{+0.19}_{-0.18}$	$0.79^{+0.10}_{-0.15}$
$M_*$	$0.68^{+0.15}_{-0.07}$	$0.76^{+0.12}_{-0.08}$	$0.53^{+0.08}_{-0.06}$
$M_{\text{BCG}}$	$-1.07^{+0.18}_{-0.08}$	$0.39^{+0.13}_{-0.08}$	$0.70^{+0.05}_{-0.05}$
$M_g$	$2.02^{+0.08}_{-0.10}$	$1.28^{+0.12}_{-0.09}$	$1.28^{+0.12}_{-0.09}$



**Fig. 15.** Posterior distributions of the normalization,  $\alpha$ , and slope,  $\beta$ , parameters. Dark and light blue regions in the two dimensional parameter planes denote  $1\sigma$  and  $2\sigma$  uncertainty, respectively. Black solid curves are the priors of the weak-lensing mass calibration of  $\alpha_{wl}$  and  $\beta_{wl}$  (Sec. 2.4.3). Black dashed lines are the resulting values.



**Fig. 16.** Posterior distributions of the intrinsic covariance of the scaling relations. Dark and light blue regions in the two dimensional parameter planes denote  $1\sigma$  and  $2\sigma$  uncertainty, respectively. Black solid curve is the priors of the weak-lensing mass calibration of  $\ln \sigma_{\text{WL}}$  (Sec. 2.4.3). Black dashed lines are the resulting values.

**Table 8.** Intrinsic covariance for the 136 XXL clusters ( $L_X$ ,  $M_*$ ,  $M_{\text{BCG}}$ , and  $M_g$ ) without the  $E(z)$  correction. Each column is the same as in Table 2.

	$L_X$	$M_*$	$M_{\text{BCG}}$	$M_g$
$L_X$	$0.79^{+0.10}_{-0.15}$	$0.08^{+0.12}_{-0.09}$	$0.10^{+0.09}_{-0.08}$	$0.30^{+0.10}_{-0.10}$
$M_*$	$0.20^{+0.22}_{-0.22}$	$0.53^{+0.08}_{-0.06}$	$0.25^{+0.06}_{-0.06}$	$0.05^{+0.09}_{-0.06}$
$M_{\text{BCG}}$	$0.18^{+0.13}_{-0.14}$	$0.68^{+0.06}_{-0.10}$	$0.70^{+0.05}_{-0.05}$	$0.04^{+0.05}_{-0.04}$
$M_g$	$0.98^{+0.01}_{-0.02}$	$0.27^{+0.28}_{-0.31}$	$0.17^{+0.15}_{-0.18}$	$0.38^{+0.08}_{-0.07}$